PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES
Volume 17 January 15, 1931 Number 1

THE PENETRATION OF 1-NAPHTHOL-2-SULPHONATE INDOPHENOL, O-CHLORO PHENOL INDOPHENOL AND O-CRESOL INDOPHENOL INTO VALONIA

By Matilda Moldenhauer Brooks

University of California

Communicated December 4, 1930

The present paper deals with a continuation of the studies on oxidation-reduction indicators[1,2,3,4] and their penetration into living plants of *Valonia*, the marine alga which has been used for most of these experiments. Previous experiments have dealt with 2-6-dibromo phenol indophenol and other indicators more negative on the E_h-pH scale as drawn by Clark.[5,6] The present series deals with indicators lying further toward the positive end of this scale. These include 1-naphthol-2-sulphonate indophenol, *o*-chloro phenol indophenol and *o*-cresol indophenol. The dyes were made by and obtained from Dr. W. M. Clark and are of special purity.*

They were dissolved in sea water in concentrations of 0.000048 M, 0.000097 M, 0.000192 M and 0.000384 M at various pH values. Sörensen's phosphate buffers were used from pH 5.8 to 8.0 and Clark's borate buffers from pH 8.0 to 9.0. The plants were placed for various lengths of time in these solutions at a temperature of 25° ± 0.5°C. At intervals plants were taken out and the sap expressed. The concentration of the dye was measured by comparison with solutions of dye of known concentrations contained in Pyrex tubes of the same size as were used for the expressed sap, and its identity was determined by means of spectrophotometric analyses. There was neither microscopic nor macroscopic evidence of injury, except as noted, nor any curtailment of subsequent survival.

The results were as follows:

It was found that 1-naphthol-2-sulphonate indophenol does not penetrate into the sap even after 24 hours' exposure of the plants at any pH value from 5.8 to 9.0. This is of interest since the same was found with the indigo sulphonates. This suggests that the sulphonate radical may be responsible for the lack of penetration of these dyes.

The remaining dyes, *o*-cresol indophenol and *o*-chloro phenol indophenol, both penetrate into the sap in a colorless form. The expressed sap can be shaken up with a trace of oxidizer such as NaOH to oxidize the dye back to its blue color so as to permit determinations of its concentration. The reduction of these dyes by the protoplasm accords with their position in the redox scale, since the more negative 2-6- dibromo phenol indophenol[1] has already been shown to enter the sap in the reduced form. The dyes which are negative to the indophenols on Clark's scale are not reduced by *Valonia*, but either do not penetrate or, like methylene blue** and Lauth's violet, penetrate as such in a colored or oxidized form into the sap.

O-cresol indophenol and *o*-chloro phenol indophenol penetrate faster from an acid solution than from an alkaline solution. However, with any one buffer salt and external dye concentration, the concentration of the dye in the sap at equilibrium is independent of the pH of the external solution. This is similar to the results obtained with methylene blue[2,3] in which the equilibrium concentration of dye in the sap was the same at all external pH values, although the rate of penetration varied with pH.

The equilibrium concentration of dye in the sap depends on the nature of the buffer salts present. Two different equilibria were established; one when phosphate buffers were used, and another when borate buffers were used. With the phosphate buffers, in every case except when the highest concentration of dye was used, the concentration of dye in the sap at equilibrium was $^1/_4$ of that of the external solution. In the remaining case it was $^1/_2$ of that of the external solution, possibly because of incipient injury, since the plants did not survive very long when kept in this concentration solution for the same period as in the other experiments. When borate buffers were used, the equilibrium concentration in the sap was only $^1/_2$ of that found with phosphate buffers. This shows definitely that the borate buffers interfere in some way with the penetration of these dyes.

In order to determine the identity of the dye which was found in the sap, spectrophotometric analyses were made of the expressed sap after the plants had been in the solution of dye for various periods of time. In the case of *o*-cresol indophenol the maximum absorption was at 610 mμ, and in the case of *o*-chloro phenol indophenol, it was at 625 mμ. These are the wave-lengths of maximum absorption for pure solutions of the dyes themselves at the pH values (>8.6) used for making the spectrophotometric determinations.[7]

This is a preliminary paper. Details will be published subsequently. Grateful acknowledgment is hereby made to the Bache Fund of the NATIONAL ACADEMY OF SCIENCES for a grant which enabled the writer to carry on this work and to the Marine Biological Laboratory at Woods Hole where the experiments were done.

* The writer wishes to thank Dr. W. M. Clark for presentation of these dyes.

** The penetration of methylene blue as such has been denied by Irwin (*J. Gen. Physiol.*, **10**, 927, 1927), who then stated that it is not methylene blue which penetrates *Valonia*, but a lower homolog of this dye. Since Irwin used an impure sample of methylene blue, the writer suggested (3) that among other causes, Irwin's results were due to this factor. In a recent paper (*J. Gen. Physiol.*, **14**, 19, 1930) Irwin concludes that her results were due to the impurity in the dye which she used.

[1] M. M. Brooks, *Amer. J. Physiol.*, **76**, No. 2, 360, 1926.

[2] M. M. Brooks, *Univ. of Calif. Publica. in Zoöl.*, **31**, 79, 1927.

[3] M. M. Brooks, *Protoplasma*, **7**, No. 1, 46, 1929.

[4] M. M. Brooks, *Proc. Soc. Exp. Biol. Med.*, **27**, 508, 1930; *Protoplasma*, **10**, 505, 1930.

[5] Clark, W. M., Cohen, B., and Gibbs, H. D., *Pub. Health Rpts.*, **40**, No. 14, 649, 1925.

[6] Clark, W. M., *The Determination of H-ions*, Williams and Wilkins, Baltimore, 1928, 317 pp.

[7] M. M. Brooks. Process of publication.

THE WORK OF THE KIDNEY IN THE PRODUCTION OF URINE

BY HENRY BORSOOK AND HOWARD M. WINEGARDEN

KERCKHOFF BIOLOGICAL LABORATORIES, CALIFORNIA INSTITUTE OF TECHNOLOGY

Read before the Academy September 23, 1930

The work performed by the kidney in the production of urine appears to be more readily susceptible of analysis by means of the laws of thermodynamics than any other complete function of the animal body. The recognition of this possibility, of course, is not new. It has attracted a number of investigators whose essays have been collected and reviewed by Cushny.[1] The purpose and method were essentially the same in all of these studies: the computation of the theoretical minimum amount of work necessary to elaborate a solution such as urine from another such as blood. None of these computations was complete; and even in the last, made in 1914,[2] a number of processes were not taken into account, viz.: the suppression of ionization of the phosphates, and other weak acids by the greater acidity of the urine, and the production of ammonia from urea by the kidney.

In the account presented below of the minimum work necessary for the production of the urine, an attempt has been made to appraise the factors omitted in previous studies; and an alternative method has been employed in the analysis of the work involved in the transport of water. It is felt that a clearer understanding of the energetics of water transport in the body in general is obtained from the alternative treatment. The minimum work has been considered as a quantity equal to the sum of the free energy changes for the transport of each constituent, including water,

from blood to urine. It has been assumed, an assumption justifiable on practical grounds, that the substances concerned may be considered as perfect solutes.

The analysis of the urine is now so nearly complete, 95 per cent of the nitrogen being accounted for, that the total work given in table 2 may be taken as the order of magnitude of the total amount of work performed by the kidney in 24 hours. The items omitted cannot add more than a very small quantity to the present tally. This prediction is permissible on account of the form of the equation by which the work is calculated.

The value obtained for the theoretical minimum amount of work necessary to produce a solution such as urine from another such as blood is quite a small quantity; in view of the total energy used by the kidney, it is surprisingly small. Possibly because the difference between the work performed and the energy used by the kidney is so large, there has been a tendency to evade explicit statement of the obvious conclusion to be drawn regarding the efficiency of the kidney considered as a machine. The following quotation from Cushny[1] is a representative expression of this attitude: "But even if the concentration of each constituent of the plasma and urine were known, and the total work were calculated according to these formulas, this would not necessarily indicate the whole energy employed in the secretion. For this measures only the energy employed in overcoming osmotic resistance, and takes no account of that entailed in the transmission of molecules of water and solid through the cells and along the tubules. Nor until more is known of the mechanism of secretion can even a general estimate be formed of the amount of energy thus used. Now if, as one school holds, the secretion of water involves actual work on the part of the cell, more energy is required for the production of the more abundant fluid, but as it is less concentrated, the work done against osmotic resistance is smaller. The total work of the kidney in producing a dilute urine may thus be greater or less than when it is more concentrated, according as the energy required to secrete water is greater or less than that required to concentrate the solids. On the other hand, if the secretion of water does not entail the loss of energy in the kidney, as is held by some authorities, an abundant secretion of dilute urine may actually involve less work than a scantier flow containing the same amount of solids."

In the foregoing quotation there is not a clear distinction made between the work which the kidney carries out, and the energy it consumes in performing this work. In view of this quotation, it may be pointed out, that on theoretical grounds no information is required regarding the mechanism by which work is effected. The theoretical minimum quantity of work, which here is equal to the change in free energy, depends only on the initial and final states; and water may be considered in exactly

the same way as the other constituents of the urine. The uncertainty regarding the mode of calculating the work of water transport is due, in part, to the employment of the cumbersome concept of osmotic pressure.

It is also easier, with the concept of free energy change than it is with osmotic pressure, to calculate directly the work involved in such processes as the suppression of ionization, and the conversion of urea to ammonia.

Calculations.—The calculation of the work performed in the production of urine is based upon the data regarding the composition of the human blood and urine given chiefly by Cushny,[1] Mathews[3] and Hawk and Bergeim.[4] In table 1 are collected only the items which comprise the work of concentration carried out by the kidney. The free energy change, $-\Delta F$, which is equivalent to the work of concentration, is calculated by means of the equation

$$-\Delta F = NRT \ln \frac{C_{\text{plasma}}}{C_{\text{urine}}} \text{ calories.}$$

The activity coefficients have been taken as unity, because the corrections, even if adequate data were available, would be negligible in view of the low efficiency of the kidney. The volume of the urine passed in 24 hours has been taken, for convenience in calculation, as 1000 cc. N is the number of mols of any one substance excreted in 24 hours, which is here equal to the molar concentration in the urine given in table 1, R the gas constant 1.987 calories per degree, and T the absolute temperature 310.1°K.

The values of $-\Delta F$ in table 1 are computed on the basis that a finite quantity of urine is produced from an infinite quantity of blood. This amount of work may for practical purposes be taken as equal to the work involved in the actual process, arterial blood⟶urine + venous blood, because the value of $-\Delta F$ is influenced much more by the value of N than by the concentration term. If, for example, we assume that most of the urea is concentrated instead of 66.6 times, 300 times, the value of $-\Delta F$ changes only from -861 calories to -1171 calories, a negligible difference in view of the low efficiency of the kidney. And even this difference would be reduced by a corresponding correction in the opposite direction for the transport of water under these conditions. Further, as Cushny states: "The passage of the blood through the kidney does not of course completely remove the impurities it contains, because only a fraction of the plasma comes into direct contact with the cells, perhaps one-fifth or less, and even if the urea, etc., is completely removed from this fraction, 80 per cent of that brought by the artery, returns in the renal vein."

Behre and Benedict[5] expressed the opinion "that no results so far available offer definite evidence of the existence of creatinine in the blood."

They believe that there is present in the blood, not creatinine, but a "chromogenic" substance which is responsible for the color with picric acid, by which blood creatinine is estimated. Nevertheless, the kidney is able to abstract creatinine from the blood and concentrate it several hundred times in the urine, as illustrated in the findings of Rehberg.[6] We have therefore, considered the urinary creatinine as arising from creatinine in the blood.

The simplest method of computing the change in state of the phosphates is to calculate from the known total concentrations of phosphate and the hydrogen-ion concentrations, in the blood and urine, respectively, the concentration of each of the ionic forms; and then by means of the equation $-\Delta F = NRT \ln \frac{C_{\text{plasma}}}{C_{\text{urine}}}$ to compute the work for the transfer of each component from blood to urine. The value of N is so small for PO_4^{---} and H_3PO_4, that these may be ignored and the whole process of concentration of phosphate with its concomitant suppression of ionization can be considered as the transfer from one set of concentrations in the blood to another in the urine, of HPO_4^{--} and $H_2PO_4^{-}$ ions. The work involved here is given in table 1.

An alternative, longer method of computing separately the work of concentration of phosphate and the suppression of its ionization can be derived from the following hypothetical process. A quantity of phosphate equal to the amount in one liter of urine, is transferred from the blood to an infinite quantity of a hypothetical auxiliary solution in which the thermodynamic environment is the same as in the plasma, except that the total concentration of phosphate is the same as in the urine. The free energy change for this step, which is the work of concentration, is -27 calories. The suppression of ionization is calculated by considering the transfer in the form of H_3PO_4, of the total amount of phosphate which is in the urine, from the concentration of H_3PO_4 in the auxiliary solution, to its concentration in the urine where, under equilibrium conditions, it is allowed to ionize. To this must be added the difference in free energy for the neutralization by OH^- ions in the urine, instead of the plasma, of the phosphate transported; and the free energy change for the transport back to the blood of the water formed in this neutralization. The sum of all these free energy changes is only approximately -11 calories.

In calculating the change in ionization of the phosphate, the hydrogen-ion concentration of the plasma was taken as 3.98×10^{-8}, and of the urine, 1×10^{-6}, the three dissociation constants of phosphoric acid in urine and the first and third constants in plasma were calculated from the data of Sendroy and Hastings,[7] corrections being made for the ionic strengths of plasma and urine which were taken as 0.13 and 0.20, respectively; for the second dissociation constant of the phosphoric acid in plasma we have

employed the value chosen by Henderson.[8] Accordingly, the values for K_1, K_2 and K_3 in the plasma were 1.2×10^{-2}, 2.3×10^{-7} and 1.4×10^{-12}, respectively; and in the urine 1.3×10^{-2}, 2.6×10^{-7} and 2.2×10^{-12}.

The free energy change due to the suppression of ionization of the bicarbonate is almost exactly balanced by the value for the free energy of neutralization. The free energy change for suppression of dissociation and neutralization of the uric acid is similarly negligible.

In the calculation of the free energy changes for calcium the total energy change is so slight that we have neglected such factors as concentrations of complex ionic and undissociated forms.

The work involved in the transport of water from the plasma to the urine was calculated in the same way as the other constituents of the urine, by the equation

$$-\Delta F = NRT \frac{N_{\text{plasma}}}{N_{\text{urine}}}, \text{ where } N \text{ denotes the mol fraction.*}$$

From the data given by Mathews, and by Hawk and Bergeim we have taken the number of mols of solids dissolved in one liter of plasma to be 0.313, and in the urine 0.765. Mathews, quoting the data of Schmidt, gives 1.0312 as the specific gravity of plasma and 901.51 gm. as the weight of the water in 1000 gm. plasma. The weight of water in one liter of plasma is therefore 929.6 gm.; the number of mols, 51.6; and the mol fraction 0.994. A liter of urine such as that postulated in table 1 contains 36 gm. of solids and is of specific gravity 1.013; the weight of the water in one liter is 977 gm.; the number of mols 54.2; and the mol fraction 0.986. From these values, $-\Delta F$ for the transport of water is $54.2\ RT \ln \frac{.994}{.986} = +267$ calories.

We have based our calculations of the free energy change for the ammonia in the urine, upon the theory of Benedict and Nash[9] that the urinary ammonia is formed by the kidney from urea; and we have assumed that this conversion is carried out with the reacting components at the concentrations at which they exist in the blood. We have preferred this view to the alternative suggestion of Bliss;[10] the methods employed by Bliss for the determination of ammonia in the blood[11],[12] must, it seems, liberate by hydrolysis of the blood proteins, amide nitrogen which is as essentially constituent of the proteins as the amino acids, and not, therefore, a factor in the acid-base equilibrium of the body.

The values employed for the standard free energies are those given by Lewis and Randall:[13] and for the solubilities, those in the International Critical Tables.

Urea ΔF_{298} (solid) = −47,280 calories.
ΔH (solid) = −78,490 calories.
ΔF_{310} (solid) = −46,020 calories.
Solubility at 37°C. = mol fraction 0.311.
Concentration in plasma = mol fraction 0.000092.
ΔF dilution = −5010 calories.
ΔF_{310} (mol fraction 0.000092) = −51,030 calories.

NH_4^+ ΔF_{298} (1.0 M.) = −18,930 calories.
ΔH = −31,700 calories.
ΔF_{310} (1.0 M.) = −18,420 calories.
ΔF dilution (to 0.000554 M. in plasma) = −4620 calories.
ΔF_{310} (0.000554 M.) = −23,040 calories.

H_2O ΔF_{298} = −56,560 calories.
ΔH = −68,270 calories.
ΔF_{310} = −56,090 calories.

H_2CO_3 ΔF_{298} CO_2 (1 Atm.) = −94,260 calories.
ΔH CO_2 = −94,300 calories.
ΔF_{310} CO_2 (1 Atm.) = −94,260 calories.
ΔF_{310} CO_2 (40 mm. Hg) = 96,070 calories.
ΔF_{310} H_2CO_3 (40 mm. Hg) = −152,200 calories.

H^+ ΔF_{310} (3.98×10^{-8} M.) = −10,500 calories.

Therefore for the reaction

$$\begin{matrix} & NH_2 \\ & / \\ C & =O \\ & \backslash \\ & NH_2 \end{matrix} + 2H^+ + 2H_2O \longrightarrow 2NH_4^+ + H_2CO_3;$$

$-\Delta F$ = +7000 calories per mol NH_4^+ formed; from which the free energy of formation of 0.0222 mols NH_4^+ from urea is +155 calories.**

The value of each of the various forms of work performed by the kidney is set out in table 2. The total amounts to −704 gm. calories. This figure is of the same order of magnitude as that obtained in previous computations. This correspondence is due to the fact that the main work of the kidney, even when all known factors are taken into account, proves to be the work of concentration.

The processes omitted in table 2 are the free energy changes for the synthesis and excretion of 0.003 mols of hippuric acid; and the excretion of the unidentified nitrogenous constituents of the urine, comprising less than 5 per cent of the total nitrogen. These unknown quantities cannot add more than a few hundred calories at the most to the total. This prediction can be made because the value of $-\Delta F$ depends mainly upon

the value of N, the number of mols transferred; when this quantity is small, the value of $-\Delta F$ is small regardless of the magnitude of the concentration change. The total in table 2 of -704 gm. calories represents, therefore, practically all the work performed by the kidney in the production of the 24-hour urine. This value is surprisingly small, approximately 0.7 gm. calories per cc. of urine per day.

TABLE 1

THE WORK OF CONCENTRATION PERFORMED BY THE HUMAN KIDNEY IN THE PRODUCTION OF THE 24-HOUR URINE

CONSTITUENT	CONCENTRATION IN PLASMA, MOLS PER LITER	CONCENTRATION IN URINE, MOLS PER LITER	WORK PERFORMED BY THE KIDNEY $-\Delta F$ GM. CALORIES
Na^+	0.135	0.152	− 11
K^+	0.00512	0.0384	− 48
Ca	0.00224	0.00375	− 1
Mg^{++}	0.00103	0.00247	− 1
Cl^-	0.104	0.166	− 48
HPO_4^{--}	0.000807	0.00326	− 3
$H_2PO_4^-$	0.000140	0.0125	− 35
SO_4^{--}	0.000312	0.0187	− 47
HCO_3^-	0.0266	0.00106	2
Creatinine	0.000089	0.00664	− 18
Urea	0.00500	0.333	−861
Uric acid	0.000238	0.00298	− 5
NH_4^+	0.000554	0.0222	− 50
		Total	−1126

TABLE 2

THE TOTAL WORK PERFORMED BY THE HUMAN KIDNEY IN THE PRODUCTION OF THE 24-HOUR URINE

KIND OF WORK PERFORMED	QUANTITY OF WORK $-\Delta F$, GM. CALORIES
Concentration	−1126
Transport of water	+ 267
Formation of ammonia from urea	+ 155
	− 704

The value is so small that we were led to consider the possibility that the kidney may perform other functions, in which much more work is performed, than in the production of urine. The high concentration of enzymes in the kidney, greater than in any other tissue except the liver, lends some plausibility to this view. But it seems to be eliminated by the observations on the oxygen consumption and heat production of the kidney. These point to the conclusion that the kidney carries out the

work of elaborating urine with an exceedingly low efficiency, in the neighborhood of one per cent. This is discussed in detail in the following communication.

Summary.—1. The work of the kidney in the excretion of urine is analyzed by means of the second law of thermodynamics.

2. It is shown that the work performed by the normal kidney in man in the excretion of urine, is of the order of magnitude of 0.7 gm. calories per cc. of urine; or 70 gm. calories per gram of nitrogen excreted.

* Barcroft employed the following equation for the computation of the work of formation of urine:

$$\text{work} = RT\,(a_u \log \frac{a_u}{a_b} - (a_u - a_b) + b_u \log \frac{b_u}{b_b} - (b_u - b_b) + \ldots.),$$

where a_b, b_b, ... and a_u, b_u, ... represent molecular concentrations of the constituents of the blood and urine. The work in the above equation is the minimum work in the formation of a liter of urine. The summarized form of this equation is

$$\text{work} = RT\left[\Sigma\,(c_u \log \frac{c_u}{c_b} + \Sigma c_b - \Sigma c_u\right].$$

Essentially this equation is

$$W = -\Delta F + \Delta(PV),$$

where $\Delta(PV) = RT(\Sigma c_b - \Sigma c_u)$, P here being the osmotic pressure, and V the volume of the solution. Barcroft estimated the pressure-volume product change from calculated osmotic pressures; the quantity obtained represents an approximate computation of the work involved in the transport of water from the blood to the urine. Stated in summarized form, this equation is somewhat more correct than the expanded form. In the expanded form presumably, only those substances in the blood are taken into account which are present in the urine. As shown below, the molecular concentrations of all the substances in the blood including those absent from the urine, and similarly the molecular concentrations of all the substances in the urine including those not in the blood, should be included in Σc_b and Σc_u, respectively. The second term in the summarized form of this equation can be shown to be equivalent to the method of calculating the work of the transport of water by the equation

$$W = NRT\,ln\,\frac{P_{\text{blood}}}{P_{\text{urine}}};$$

where P represents the vapor pressure of water.

We are indebted to Professor R. G. Dickinson and to Professor R. C. Tolman for this derivation of the relationship between the two equations.

$$-\Delta F = NRT\,ln\,\frac{N_{\text{blood}}}{N_{\text{urine}}};$$

which is an approximation of the rigorous equation

$$-\Delta F = NRT\,ln\,\frac{P_{\text{blood}}}{P_{\text{urine}}};$$

where P_{blood} and P_{urine} represent the vapor pressures of water in blood and urine,

respectively. The vapor pressure, by Raoult's law, is proportional to the mol fraction of the solvent.

$$\therefore P_{\text{blood}} = K\,(I - x_a^b - x_b^b - x_c^b - \ldots\ldots);$$

$$\text{and} \quad P_{\text{urine}} = K\,(I - x_a^u - x_b^u - x_c^u - \ldots\ldots);$$

where x_a^b, x_b^b, x_c^b, and x_a^u, x_b^u, x_c^u, represent the mol fractions of the constituents other than water a, b, c, and a', b', c', in the blood and urine, respectively.

$$\therefore -\Delta F = NRT \ln (1 - \Sigma x^b) - NRT \ln (1 - \Sigma x^u).$$

The value of N is the number of mols of water transported and may be designated as $N^u_{H_2O}$.

The value of the expression log (1 + x), when x^2 is less than 1, is $x - {}^1/_2x^2 + x^3 - {}^1/_4x^4 \ldots\ldots$ In plasma the value of x is 0.007 and in urine 0.015. We may, therefore, in both cases, ignore all terms but the first, so that the value of $-\Delta F$ becomes

$$NRT\Sigma x^u - NRT\Sigma x^b.$$

Since by definition the mol fraction is $\dfrac{n_1}{n_1 + n_2 + n_3 + \ldots\ldots}$, where n_1 and n_2 are very small compared with n_3, which we shall take to represent the number of mols of water in a liter of solution, therefore, approximately

$$\Sigma x^b = \frac{\Sigma c^b}{n^b_{H_2O}},$$

where Σc^b is the sum of the concentrations of all the substances dissolved in a liter of blood, and $n^b_{H_2O}$ is the number of mols of water in a liter of blood.

$$\text{Similarly } \Sigma x^u = \frac{\Sigma c^u}{n^u_{H_2O}}.$$

$$\because N^u_{H_2O} = n^u_{H_2O} \text{ and is approximately equal to } n^b_{H_2O}.$$

$$\therefore N^u_{H_2O}\,\Sigma x^b = \Sigma c^b; \text{ and}$$

$$N^u_{H_2O}\,\Sigma x^u = \Sigma c^u.$$

$\therefore -\Delta F = RT\Sigma c^u - RT\Sigma c^b$, which, allowing for the conventional difference in sign, is the term in the summarized form of Hill's equation for the transport of water.

The following is a more rigorous thermodynamic proof. Taking $\bar{F}$, f, N, $\bar{V}$, P to represent respectively partial molal free energy, fugacity, mol fraction, partial molal volume, and the applied external pressure, and assuming throughout perfect solutions, as is the case in the Barcroft equation—

$$\left(\frac{\partial \ln f}{\partial \bar{F}}\right)_{P,\,T} = \frac{1}{RT}; \tag{1}$$

$$\therefore \bar{F} = RT \ln f + C;$$

$$\therefore \bar{F} - F' = RT \ln \frac{f}{f'}. \tag{2}$$

$$\left(\frac{\partial ln\, f}{\partial P}\right)_{T,\,N} = \frac{\overline{V}}{RT}; \tag{3}$$

$$\therefore\ RT\int_{f'}^{f} d\ ln\, f = \int_{P}^{P'} \overline{V}\, d\, P;\ \text{where } P'$$

is the pressure that must be applied to the solution in the state where fugacity is f' to change its fugacity to f;

$$\therefore\ RT\ ln\ \frac{f}{f'} = \overline{V}\,(P' - P). \tag{4}$$

$$f = f^{\circ}N; \tag{5}$$

where f° is the fugacity of the substance, here water, in the pure state, and is a constant,

$$\therefore\ ln\, f = ln\, f^{\circ} + ln\, N;$$

$$\therefore\ \int d\ ln\, f = \int d\ ln\, N + C';$$

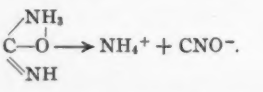
$$\therefore\ ln\ \frac{f}{f'} = ln\ \frac{N}{N'}; \tag{6}$$

$$\therefore\ \overline{F} - F' = RT\ ln\ \frac{f}{f'} = RT\ ln\ \frac{N}{N'} = \overline{V}\,(P' - P). \tag{7}$$

$$\overline{V}\,(P' - P) = RT\,(\Sigma C' - \Sigma C), \tag{8}$$

where ΣC is the total molar concentration of the dissolved substances.

** If the mechanism of the formation of ammonia from urea is taken to be that suggested by Werner,[14] the reaction may be written as follows:



Assuming the concentration of the cyanate ion to be of the order of magnitude of 0.0002 molar, and correcting the standard free energy of the cyanate ion for the change in temperature from 25 °C. to 37 °C., $-\Delta F$ for the reaction in the blood is +600 calories per mol of ammonium ion formed. Therefore, for 0.022 mols of ammonium ion the free energy change is +13 calories.

Since the formation of ammonia from urea appears to be a function only of the healthy kidney, it is open to question if the free energy change for this reaction, with its positive sign, is legitimately considered, as it is here, as work done upon the kidney.

THE ENERGY COST OF THE EXCRETION OF URINE

BY HENRY BORSOOK AND HOWARD M. WINEGARDEN

KERCKHOFF BIOLOGICAL LABORATORIES, CALIFORNIA INSTITUTE OF TECHNOLOGY

Read before the Academy, September 23, 1930

It has been shown that the work performed by the normal kidney in man in the excretion of urine is of the order of magnitude of 0.7 gram calories per cc. of urine; or 70 gram calories per gram of nitrogen excreted. It is instructive to compare the quantity of work performed by the kidney with the energy consumption of the kidney during the performance of this work.

The energy consumption of the kidney has been estimated by a number of experimenters and by different methods; and the values obtained are all of the same order of magnitude. The concordance of these values justifies an estimate of the efficiency of the kidney, i.e., of the ratio of the work performed, calculated from the constitution of the urine, to the energy used. The efficiency of the kidney, defined in this way, appears, even in health, to be about 1–2 per cent. So far as the authors are aware, there are no data inconsistent with this low figure. Rather the reverse, this association of the thermodynamic work with the observed energy consumption of the kidney permits the correlation of a large number of facts regarding the behavior of the kidney in health and disease.

The earliest measurements of oxygen consumption of the kidney are those of Barcroft and Brodie[15,16], and of Barcroft and Straub.[17] These show that the metabolism of the kidney is increased when it is performing work, i.e., when it is producing a urine in which the dissolved substances are present in greater concentrations than in the plasma. Barcroft and Straub found, for example, that the excretion of large quantities of chloride in the urine, in concentrations nearly the same as that in the plasma, was not accompanied by an increase in the consumption of oxygen; while the excretion of sulfate or of urea in the urine, in much greater concentration than in the plasma, was attended by large increases in the oxygen consumption. These results might have been predicted from thermodynamic theory. The association of increased oxygen consumption only with increased work, is strikingly exhibited in the behavior of the kidney damaged in various ways, as by caffeine, mercuric chloride, bleeding, or dilution of the blood or plasma with saline. Under these conditions the oxygen consumption of the kidney was not increased when diuresis was induced by sulfate or urea, and at the same time these substances were not concentrated in the urine as they were by the undamaged kidney, but were passed in nearly the same concentrations as in the plasma. In one case a larger quantity of sulfate was excreted at a low concentration,

without increased oxygen consumption, than in another where there was a large increase in the oxygen consumption, and a high concentration of the sulfate in the urine. The observations of Barcroft and Straub have been confirmed by Knowlton and Silverman[18] and by Bainbridge and Evans[19] (in one case on the isolated kidney). Winfield[20] found that the oxygen consumption of the kidney fell to $^1/_4$–$^1/_5$ of its normal value when the urine became isotonic after the intravenous injection of saline.

The dependence of the concentrating function of the kidney on concomitant oxidations is strikingly illustrated in the effects of cyanide on the kidney. Starling and Verney[21] observed, after the addition of KCN to the circulating blood in a heart-lung-kidney preparation, that the NaCl concentration in the urine rose to that in the plasma, i.e., the "concentration" of water fell to that of the serum, while the concentrations of urea and of sulfate fell markedly. On changing the circulating blood to one free from KCN, the chloride concentration fell again to a value not far from the original level, while the urea and sulfate concentrations rose a little, the sulfate more than the urea, toward their former levels. Similarly the power of the kidney to excrete and concentrate phenol red was also markedly reduced after poisoning with cyanide. The observations of Marshall and Crane[22] are essentially the same. These observers found that the urine contained approximately the same amount of water chloride, and bicarbonate after a period of temporary asphyxia induced by occlusion of the renal artery for 20 minutes, but much less urea, phosphate sulfate, creatinine and ammonia.

Three groups of workers claim to have been unable to confirm the observations described above. The first apparently contradictory results were those obtained by Tamura and Miwa.[23] It is not possible to interpret their results, or to compare them with those of Barcroft and Straub, first, because the urine analyses are not given. For example, if the excretion of urea is markedly reduced during a sulfate diuresis, it is possible that in spite of an increased excretion of sulfate, the total work performed by the kidney may be reduced; or a great change in the urinary chloride concentration may change the sign of the free energy term for the excretion of water. Tamura and Miwa found, in the rabbit, that diuresis induced by sulfate and caffeine, did not increase the oxygen consumption, while diuretin and salicylates provoked large increases. The employment of different animals by the two groups of experimenters, the rabbit in the experiments of Tamura and Miwa, the cat and dog in the experiments quoted here performed by Barcroft and Brodie, and Barcroft and Straub, is the second condition which prevents a comparison of the results obtained. The following typical observations show the distinct differences in function between the rabbit kidney on one hand, and the kidney of carnivora or omnivora on the other. After the injection of massive doses, the data of

Mayrs[45] show in the rabbit urea concentrated in the urine to only three times its concentration in the plasma, sulfate six times, phosphate four to five times and creatinine four times. On the other hand, in the urine of the dog, cat or man, urea may be concentrated sixty times, sulfate one hundred times, phosphate twenty times, and creatinine one to two hundred times the concentration in the blood. The experiments of Frandsen[24] show that the excretion of ingested sodium chloride or urea is very slow in the normal rabbit. Frandsen found in two experiments an average increased hourly excretion of only 0.016 and 0.017 gm. of nitrogen, respectively, in the four-hour period after the ingestion of 0.75 gm. of urea in addition to the usual ration. In another experiment in which a rabbit received 1.0 gm. of urea 3 hours after the routine meal, there was a uniform increased excretion of 0.052 gm. of nitrogen during the next four hours. It is possible, therefore, that in spite of the diuresis observed, the amount of the injected substances actually excreted in the experiments of Tamura and Miwa was not greatly increased. In view of the relatively slight degree to which substances are concentrated in the urine of the rabbit it seems probable, further, that the formation of urine by the rabbit's kidney entails considerably less work than is the case in carnivora or omnivora; and accordingly, the experimental error in the detection of differences is magnified.

Recently Hayman and Schmidt[25] described experiments in which "the injection of urea, caffeine, sodium sulfate, or certain drugs was not followed by any characteristic effect on apparent oxygen metabolism." Their protocols show that in four experiments out of six, caffeine was administered 10–20 minutes before sodium sulfate; here sodium sulfate did not elicit any increased consumption of oxygen, even though the flow of urine was increased. In one of the remaining experiments, in which the injection of sulfate was not preceded by caffeine, the protocol gives the largest value for oxygen consumption recorded in their experiments. The observations of Hayman and Schmidt, therefore, as described in their protocols, confirm the observations of Barcroft and Straub, "that caffeine virtually abolished the specific effect of sodium sulfate," of increasing the oxygen consumption of the kidney. Furthermore the experimental procedure of Hayman and Schmidt was rather drastic. In one experiment within a period of two hours after the completion of the operation, there were injected 150 cc. blood, 165 cc. saline, 80 mgm. caffeine, 20 cc. of 20 per cent sodium sulfate, pituitrin, adrenalin, and morphine. For this reason it is difficult to interpret the observations of Hayman and Schmidt. Hardly enough time, it seems, was allowed for the effect of one of the injected substances to manifest itself, before another was injected; much longer time was allowed in the experiments of Barcroft and Straub. In addition the data of the urine analyses are not given.

Fee[26] (1929), using a heart-lung-kidney preparation, obtained results which were taken to confirm the negative results reported by Hayman and Schmidt. Fee's statement is that "the actual excretion of large amounts of sulfate by the kidney is not attended with an increase in the renal oxygen consumption unless there is a concomitant increase in renal blood flow." It is questionable if the results of experiments on a heart-lung-kidney preparation may be compared with those obtained by Barcroft and Straub with the kidney *in situ*. Starling and Verney, and Fee and Hemingway[27] found that the urine from the heart-lung-kidney preparation is hypotonic, while the healthy mammalian kidney *in situ*, of course, excretes hypertonic urine. The production of either hypertonic or hypotonic urine is attended with an increased oxygen consumption; when the hypotonicity is reduced in the isolated kidney, by pituitrin for example, the oxygen consumption is also decreased. It is possible that the injection of sulfate under these conditions has the same effect as pituitrin, i.e., increases the concentration of salt in the hypotonic urine, rendering it more nearly isotonic, and thereby reduces the work and the oxygen consumption of the kidney. It is difficult to obtain a clear interpretation of the results reported by Fee, because the analyses of the urine, in detail are not given. The analysis of the work of secretion of hypotonic urine, below, shows that under the conditions of this experiment the result to be expected of an increased concentration of sulfate, from thermodynamic considerations, may be a reduced rather than an increased consumption of oxygen, due to a large decrease in the work of excretion of water. The results as given, therefore, may be taken to confirm, rather than to disagree with the hypothesis that increased work by the kidney is attended by a demonstrably increased consumption of oxygen. It is not certain, moreover, that the kidney in these experiments was in a "normal" state; the experiment reported in detail "was terminated by oedema" within less than an hour after the injection of the sulfate. More weight may therefore be placed on the observations on the effects of cyanide by Starling and Verney on the same kidney preparation as used by Fee.

Further confirmation of this hypothesis, that work of concentration entails increased metabolism in the kidney, is contained in the observation by Barcroft and Brodie of a large increase in renal oxygen consumption during phlorhizin glycosuria. In phlorhizin glycosuria the concentration of sugar in the urine may become very high, one hundred times that in the plasma; while the total nitrogen, urea, phosphates, and chlorides show no significant change. The observation of an increased oxygen consumption when sugar is excreted in highly concentrated solution is significant here, because the severely damaged kidney excretes approximately an isotonic urine, containing sugar in the same concentration as in the plasma; and its oxygen consumption is markedly reduced.[21]

The production of hypotonic urine, can be shown by the same method as that employed in the previous communication to entail work on the part of the kidney; and may, therefore, be expected to be attended by an increased consumption of oxygen. Only the excretion of a strictly isotonic urine, intermediate between the extremes of hyper- and hypotonicity, entails no work by the kidney. This can be illustrated by assuming an extreme case, a "urine" consisting only of a 1 per cent solution of urea; in which the mol fraction of water, therefore, is approximately 0.997. The free energy change for the transport of the water in a liter of such "urine" is now negative instead of positive, amounting to about −100 calories; $-\Delta F$ for urea is −430 calories. The total of −530 calories is of the same order of magnitude as the total free energy for the production of one liter of average hypertonic urine. This is due to the fact that a relatively large quantity of water is transported from a lower concentration in the plasma to a higher concentration in the urine. Further, the conditions under which such a urine is produced are usually those of diuresis so that the value of N for water especially is greatly increased, with the result that the work of the kidney per unit time is quite large. The tonicity of the urine depends mainly upon the sodium chloride and the urea; when the urea excretion is constant and its concentration in the urine is less than the total molecular concentration of the plasma, the concentration of sodium chloride as a rule determines whether or not work is done in the transport of water from the plasma to the urine. When the chloride concentration is low, so that the mol fraction of water in the urine is greater than in the plasma, the kidney performs work in transporting water from plasma to urine. On the other hand, when the concentration of chloride in the urine is greater than in the plasma, so that the mol fraction of water is less in the urine than in the plasma, water is "diluted" in passing from plasma to urine, and $-\Delta F$ becomes positive for the transport of water, i.e., the kidney "gains" work.

The physiological evidence bearing upon this aspect of renal function, the work of secretion of hypotonic urine, is not abundant; but as far as it goes it is completely confirmatory of the above thermodynamic predictions. The urine of the frog is always hypotonic to the plasma; and the concentration of nitrogen is practically the same as the non-protein nitrogen in the blood.[28] When the frog's kidney is poisoned with caffeine, or corrosive sublimate, the concentration of salts in the urine approximates that in Ringer's solution. After the administration of cyanide "the fluid issuing from the ureter comes to resemble the glomerular filtrate, sugar tends to appear in the urine, the chloride is less reduced than usual, potassium and calcium tend to approach the concentration of the perfusing fluid, and urea is no longer concentrated."[1] Narcotics, which Keilin[29]

recently has shown are selective poisons for one type of widely distributed oxidation mechanism, have the same effect as cyanide.

The isolated mammalian kidney, perfused by the heart-lung preparation as mentioned above, resembles a frog's kidney in that the urine secreted is hypotonic with respect to salts, although urea and sulfate are concentrated to a considerable extent. Starling and Verney found that this kidney after being poisoned with cyanide passed a urine which was isotonic with respect to the serum in every detail. Similarly, Fee and Hemingway working with the same heart-lung-kidney preparation found that the oxygen consumption increased as the urine contained less and less chloride. The significance of this observation is augmented by the fact that the oxygen consumption of the damaged or diseased kidney is less than normal.[17,39,40] When pitruitrin was added to the perfusing fluid in the experiments of Fee and Hemingway[27] the chloride concentration rose, and at the same time the oxygen consumption became less.

The evidence, therefore, seems clear that the performance of what may be called thermodynamic work by the kidney involves an expenditure of energy which is derived from oxidations within the renal cells. It is instructive, therefore, to compare the work performed by the kidney with the energy used, i.e., to determine the "efficiency" of the kidney.

The discussion of the relationship between the work performed and the energy used by the kidney is facilitated by the following explicit definition of the terms employed and the assumptions made.*

W = Minimum work necessary to produce 1000 cc. *urine* from infinite quantity of *blood* of character entering kidney.

W_1 = Minimum work necessary to produce 1000 cc. *urine* + venous blood from the actual quantity of *blood*.

$W_1 \approx W$. The justification for this assumption has been discussed in the previous communication.

Q = Heat of combustion of fuel burned while 1000 cc. urine are excreted.

Q_1 = Maximum work available from this combustion. In a previous communication[46] it was shown that the free energy of combustion of fat and of glucose in the body was, in each case, nearly equal to the heat of combustion. No data are available regarding the free energy of combustion of the proteins. But is probable that no error is involved in assuming that $Q_1 \approx Q$.

Q_m = Heat of combustion of fuel used in maintaining the kidney alive for the length of time necessary to produce 1000 cc. of urine but not connected with the production of urine.

Q_{iso} = Additional heat of combustion required to produce 1000 cc. urine of such character that $W_1 \approx W = 0$. For practical

purposes this may be taken to represent the additional heat produced by the kidney in the formation of 1000 cc. isotonic urine.

The efficiency, E, is a conventional term, which may be defined in a number of ways. Three possibilities are: $E = \frac{W}{Q}$; $E' = \frac{\Delta W}{\Delta Q}$, where Δ is the change accompanying any change in the blood and urine; and $E'' = \frac{W}{Q - (Q_m + Q_{iso})}$. It is obvious that only E' may be expected to show any degree of constancy with variation in the amount of work performed, i.e., with large variations of ΔW.

The values for the rate of metabolism of the kidney determined by a number of different methods are remarkably concordant: Barcroft and Brodie, and Barcroft and Straub observed rates of oxygen consumption per gram of "resting" kidney per minute, ranging from 0.008 to 0.075 cc. in the dog, from 0.03 to 0.12 cc. in the cat, and from 0.046 to 0.13 cc. in the rabbit. Fee and Hemingway obtained similar values for the isolated kidney (presumably of the dog) of 0.03 to 0.2 cc. per gram per minute. The term "resting" value, is to some extent a misnomer here, because the kidney even at rest is producing urine; and it is possible that the bulk of the urine excreted in a day is elaborated by the "resting" kidney i.e., not in a state of diuresis. For the purpose of estimating the order of magnitude of renal efficiency in man, we have taken as an average value for the rate of oxygen consumption of the kidney over 24 hours the relatively low figure of 0.03 cc. per gram per minute; and an average weight for the human kidney of 150 grams.[30] On this basis the renal oxygen consumption in man in 24 hours is of the order of magnitude of 13 liters. Assuming an R. Q. of 0.80, the caloric equivalent of this quantity of oxygen is 62 kg. calories.[31] Since the total work performed by the kidney as set out in table 2 of the previous communication is 704 gram calories, the "efficiency" of the kidney, i.e., W/Q, is approximately 1 per cent. Barcroft and Brodie calculated from their experiments ratios of work performed to energy consumed during periods of diuresis, i.e., $\Delta W/\Delta Q$, and obtained values for urea excretion of 1.75 and 0.14 per cent; it should be mentioned that, because these values for renal efficiency are so low, Barcroft and Brodie hesitated to accept them as correct. Nevertheless, as pointed out before, the total of 704 gram calories represents probably all but a negligible quantity of the work performed by the kidney, and the values employed for oxygen consumption are low rather than high. During diuresis Barcroft and Brodie found that the oxygen consumption of the kidney may be relatively 5–10 times greater than the oxygen consumption of the body as a whole.

Janssen and Rein[32] recently attempted to determine the metabolism of the kidney by measuring the difference in temperature between the

arterial and venous blood. They obtained values for the metabolism of the kidney of a similar order of magnitude as those above.

In view of these low calculated values of renal efficiency, it was desirable to obtain, if possible, some estimate of renal efficiency in man. A number of such experiments are on record in the literature; but the results given are so discordant that it was decided to carry out the experiments again.

The experiments described below were performed on young men. These came into the hospital immediately after the evening meal, between 7 and 8 P.M., on the night before the experiment. From then until the end of the experiment at 11–12 A.M. of the next day, they remained in bed and received no food or drink, except the urea solution or water as indicated in table 1. The subjects voided during the night, the following morning on waking, and then during the experiment at the times indicated. The metabolism was determined by measuring the oxygen consumption over a period of five or six minutes by means of the McKesson "Metabolor." In order to accustom the subjects to the procedure of determining the metabolic rate, a number of determinations were made, after the subject was in bed the night before. The practice during the experiment was to obtain, at intervals of one-half hour, three consecutive values for the metabolism before administering the urea solution or water; and then for four or six half-hourly intervals afterward. Urine was collected immediately after each estimation of the metabolism. The subjects experienced no discomfort, and only brief transient sensations of any sort after the ingestion of the urea. In order to minimize the effects of water *per se*, the urea was given in very concentrated solution. Since, according to Dubois[38] it is not possible to maintain a human subject in a steady basal state for more than a few hours, the experiments as a rule began at 8:00 A.M., and all values for the metabolism taken after 12:00 noon have not been used. The urea solution was warmed to 37°C. before it was given to the subject to drink. Chlorides were determined by the Volhard-Arnold method,[4] urea, with the use of urease.

The results obtained are summarized in table 1. They show an unequivocal rise in metabolism accompanying increased excretion of urea. The calculated values given in the last column of table 1 indicate the low order of magnitude of $\Delta W/\Delta Q$. This ratio is a better measure of the working efficiency of the kidney than either of the other two efficiency ratios. The fact that the value of this ratio is of the same order of magnitude as W/Q and $W/\Delta Q$ obtained from estimations of the oxygen consumption of the isolated kidney, indicates that the sum of Q_m and Q_{iso} probably is small compared with the additional energy expended in the performance of work by the kidney. This is substantiated by the findings of Winfield, and by the oxygen consumption of the poisoned as compared with the unpoisoned kidney.

The data are not sufficiently comprehensive, however, to throw any light on the problem of the variation of the value of the ratio $\Delta W/\Delta Q$ with the quantity of work performed. It seems probable that this quotient is not a constant, but that it will vary with the amount of work performed and with the condition of the kidney, and of the animal as a whole. On the other hand, the value of two per cent for $\Delta W/\Delta Q$ seems to be a maximal rather than a minimal value.

It is obvious that the explanation for the low efficiency of renal function is to be sought in the mechanism of the production of urine. It is possible that the reabsorption of large quantities of water, according to Rehberg several hundred times the volume of the urine, with the corresponding amounts of sodium chloride and sugar, may be the process which incurs a large expenditure of energy by the kidney, and for which there is nothing to show in the urine. Nevertheless, the figures computed for the work of reabsorption of even 200 liters of water, with the corresponding amount of electrolytes and sugar, during the formation of 1000 cc. of urine, calculated as in table 1 in the previous communication, amount to only a few per cent of the energy used. It would appear that the work of such a hypothetical reabsorption is performed with a low efficiency similar to that obtained for the final formation of urine.

A low mechanical efficiency appears to be characteristic of glandular function. For example, the data quoted in Bayliss[47] give 1.5 mg. sugar, or six gram calories, as the amount of energy consumed in the secretion of one cc. of saliva, and the calculated work for the formation of saliva from blood is much less than that for urine.

These are not the first observations of an increase in metabolism following the ingestion of urea. Zuntz,[34] quoting from experiments carried out by Steck, records that in dogs there was a rise in metabolism of 0.49–0.53 calories per gm. of nitrogen ingested and in man 0.94–1.37 calories. The rise in metabolism occurred only during the first two hours, and in man only one-sixth of the ingested urea was excreted. If the values obtained by Steck in men are accordingly multiplied by six the results are the same as those recorded in table 1. Recently Lublin[35] measured the change in metabolism in men following the ingestion of large quantities of urea solution, and observed a rise in the metabolic rate in one case of 8 per cent, and in another of 12 per cent of the basal rate; the high point coinciding with the time of greatest nitrogen excretion, and not of greatest urine volume.

A rise in metabolism following the ingestion of urea has been observed also by Lusk[36] and by Grafe,[37] on whose authority, nevertheless, the conclusion has been accepted generally that the excretion of urea is not accompanied by any demonstrable increase in metabolism. This conclusion has passed unchallenged in spite of the work on the oxygen consumption

TABLE 1

THE INCREASE IN METABOLISM IN MEN DURING THE EXCRETION OF UREA

NAME	SUBJECT: BODY SURFACE (SQUARE METERS)	SUBJECT: YEARS, AGE	TIME: OF INGESTION OF UREA SOLUTION OR OF WATER	TIME: OF DETERMINATION OF METABOLISM	TIME: OF URINE PRODUCTION	VOLUME OF URINE (CC.)	EXCRETION OF CHLORIDE: CONCENTRATION AS NaCl (GM./100 CC.)	EXCRETION OF CHLORIDE: AMOUNT PER HOUR AS NaCl (GM./HR.)	EXCRETION OF UREA: CONCENTRATION (GM./100 CC.)	EXCRETION OF UREA: AMOUNT PER HOUR (GM./HR.)	METABOLISM, KG. CAL. PER SQUARE METER PER HOUR (KG. CAL.)	INCREASE IN METABOLISM FOR ADDITIONAL NITROGEN EXCRETION, KG. CAL./GM. (KG. CAL.)
R. L. B.	1.77	27	at 9.38,		8.02– 8.32	25	1.08	0.54				
			37.5 cc.	8.55– 9.00	8.32– 9.02	26	1.06	0.55	0.80	0.42	36.3	
			of 40	9.25– 9.30	9.02– 9.34	30	0.98	0.55	0.78	0.44	36.4	
			per cent	9.55–10.00	9.34–10.02	35	0.82	0.83	0.96	0.97	39.8	
			Urea	10.25–10.30	10.02–10.33	64					39.0	10
			solution	10.55–11.00	10.33–11.05	41	0.94	0.69	1.19	0.87	38.0	
				11.35–11.40	11.05–11.42	43					41.3	
B. B.	1.58	26	at 9.36,	8.00– 8.05							37.5	
			37.5 cc.	8.25– 8.30	8.06– 8.31	30	1.26	0.91	0.77	0.55	36.7	
			of 40	8.55– 9.00	8.31– 9.03	27	1.24	0.63	0.65	0.33	..	
			per cent	9.25– 9.31	9.03– 9.32	28	1.24	0.72	0.60	0.35	37.3	
			Urea	9.55–10.00	9.32–10.02	31	1.08	0.74	0.81	0.55	38.6	11
			solution	10.25–10.31	10.02–10.31	36					40.0	
				10.55–11.00							40.0	
				11.05–11.12	10.31–11.05	39	0.90	0.68	1.10	0.83	39.7	
				11.25–11.30	11.05–11.32	38					41.3	
G. H. P.	1.90	19	at 7.50,		6.43– 7.49	46	1.39	0.58	0.96	0.40		
			20 gm.	7.10– 7.15							41.8	
			Urea	7.40– 7.45							42.4	
			in 35 cc.	8.10– 8.15	7.49– 8.23	39	1.18	0.81	0.73	0.50	44.0	
			water	8.45– 8.50	8.23– 8.52	40	1.00	0.83	1.08	0.90	44.0	7
				9.00– 9.05							44.1	
				9.20– 9.25	8.52– 9.32	67	0.94	0.94	1.22	1.22	43.6	

R. C. H.	1.77	21 at 8.59, 37.5 cc. of 40 per cent Urea solution	8.15– 8.22	7.53– 8.54	34.5	1.08	0.37	0.83	0.28	39.0	
			8.45– 8.52							38.5	
			9.15– 9.20	8.54– 9.27	22.5	1.02	0.42	0.87	0.36	39.0	
			9.45– 9.55	9.27– 9.56	28.0	0.79	0.46	1.16	0.68	41.5	
			10.15–10.21	9.56–10.22	25.5					40.3	6
			10.45–10.52	10.22–10.53	28.0	0.74	0.36	1.38	0.67	38.9	
			11.15–11.22	10.53–11.24	22.5					40.3	
O. P. J.	1.99	21 at 9.47, 50 cc. of 40 per cent Urea solution	8.00– 8.06	6.30– 9.26	164	0.96	0.54	0.66	0.37	38.3	
			8.50– 8.55							38.3	
			9.45– 9.50	9.26– 9.47	20	0.96	0.55	0.66	0.38	38.3	
			10.00–10.06	9.47–10.06	122	0.70	2.70	0.48	1.85	47.1	
			10.30–10.36	10.06–10.38	132	0.56	1.39	0.47	1.16	45.8	
			11.00–11.06	10.38–11.07	85	0.68	1.20	0.76	1.34	43.4	11
			11.30–11.36	11.07–11.39	54	0.89	0.90	1.13	1.14	38.2	
			12.00–12.15	11.39–12.45	73	1.03	0.68	1.36	0.90	39.7	
			12.37–12.44							40.0	
O. P. J.	1.99	21 at 8.06, 35 cc. water	7.05– 7.12	7.18– 8.06	65	1.46	1.19	0.59	0.48	38.2	
			7.30– 7.35							37.2	
			7.55– 8.00							38.4	
			8.25– 8.30	8.06– 8.35	59	0.93	1.14	0.40	0.49	39.2	
			8.55– 9.00	8.35– 9.05	73	0.73	1.07	0.31	0.45	38.7	
			9.25– 9.30	9.05– 9.35	57	0.98	1.12	0.42	0.48	38.1	
R. D. M.	1.84	18 at 8.57, 30 cc. water	8.15– 8.20							40.8	
			8.45– 8.50	8.23– 8.54	15	1.39	0.40	0.67	0.19	40.3	
			9.15– 9.20	8.54– 9.24	18	1.43	0.51	0.68	0.24	39.7	
			9.45– 9.51	9.24– 9.53	20	1.56	0.63	0.62	0.25	38.3	
			10.15–10.20	9.53–10.23	20					38.7	
J.	1.66	18 at 9.02, 35 cc. water	7.55– 8.00	7.34– 8.04	23					42.7	
			8.20– 8.25	8.04– 8.31	16.5	1.34	0.53	1.00	0.39	44.8	
			8.50– 8.55	8.31– 9.00	17					43.6	
			9.30– 9.35	9.00– 9.39	20					43.1	
			9.55–10.00	9.39–10.04	10	1.38	0.39	1.17	0.33	38.4	
			10.20–10.25	10.04–10.31	13					43.2	

of the isolated kidney, and in spite of the fact that the conclusions both of Lusk and of Grafe rest in each case upon only two experiments, of which the results do not agree with each other.

The experiments of Lusk were carried out on dogs. In one, 17 grams of urea dissolved in 150 cc. of water were given at 12:00 noon. At 12:45 the dog vomited. Nevertheless, a large quantity of urea must have been absorbed, because the nitrogen elimination was increased from 0.154 to 0.424 gm. per hour after the administration of the urea. The basal metabolism of this dog varied from 14.7 to 18.3 calories; the general average given is 16.2 calories. In the interval designated 1–2 (which presumably was the second hour after administration of the urea) the metabolism rose to 20.42 calories per hour, and in the third hour was 18.8 calories per hour. The results of this experiment were rejected by Lusk because the animal vomited during the first hour after administration of the urea. Benedict and Emmes[38] have shown that intestinal movements have a negligible effect on the metabolic rate; and the rise in metabolism observed here occurred over a period of two and a quarter hours after the vomiting. It is important that an increase in metabolism in this experiment was obtained both by direct and indirect calorimetry. In the second experiment a dog was given 12 gm. urea in 150 cc. water at noon, and the recorded observations on the metabolic rate were made during the third and fourth hours after the administration of the urea. No increase in metabolism was observed in this interval. On this one result is based the conclusion that the excretion of urea is not accompanied by an increase in metabolism. Our observations on human beings, and those of Zuntz, indicate that the greatest excretion of urea generally occurs in the first two hours after the ingestion of urea; after this time the excretion of urea may be little more than in the pre-administration period. It was precisely in this last period that Lusk failed to observe any increase in metabolism; the large increase observed in the first experiment, on the other hand, was in the second hour. These results are not, therefore, antithetical to those obtained by Steck, Lublin, or those given in table 1.

In Grafe's experiments, both carried out on the same human subject, 22 gm. urea were administered in 750 cc. bouillon. In one case there was a decline of 2 per cent over a period of six hours; and in the other, over a period of $3^1/_2$ hours there was a rise in metabolism of 8 per cent. Apart from the discordance of the results, these experiments of Grafe are open to the criticism, based upon the observations of Dubois, that it is not possible to maintain a human subject in a steady basal state for more than three or four hours, and results obtained after such a period cannot be considered as reliable.

The question of the energy cost of the elimination of nitrogen is obviously an important one in the elucidation of the specific dynamic action

of protein. The experiments of Lusk and of Grafe were carried out ostensibly for this purpose, and their conclusion seems justified, in spite of the discrepancies in their data, that the whole of the specific dynamic action of protein cannot be due to the work of the kidney in the elimination of nitrogen. On the other hand, their conclusion that the excretion of nitrogen by the kidney incurs no expenditure of energy is not warranted by their own data. Moreover, it is improbable from the observations on the oxygen consumption of the isolated kidney, and it is contradicted by the observations of Steck, of Lublin, and those summarized in table 1. The significance of the work of the kidney in the specific dynamic action of protein is discussed elsewhere.

The conception of the normal kidney as a chemical machine which possesses a great working capacity, but which performs its work with a very low efficiency readily lends itself to the explanation of the effects of disease of the kidney on the composition of the urine and blood.

In general there may be said to be two possible effects of disease on the kidney: one, that the diseased kidney tends to perform its normal quantity of work, but with a lower efficiency; the other, that the effect of disease is to reduce the capacity of the kidney for work. It may be expected *a priori*, since "loss of function" is one of the classical signs of disease, that the second rather than the first of the two alternative reactions to injury is the case in the kidney. The observations of Barcroft and Straub mentioned above, to the effect that caffeine, mercuric chloride, and anaemia reduce the oxygen consumption of the kidney, and at the same time change the composition of the urine so that it approximates the plasma, support this surmise. Its correctness is sustained further by the findings of Tribe, Hopkins and Barcroft,[39] and of Tribe, Harvey and Barcroft,[40] that the oxygen consumption of the kidney damaged by uranium or by diphtheria toxin was always lower, and was frequently less than half the average value obtained for the normal. The effects of cyanide and of narcotics also support this view.

The human kidney when it is diseased fails in a like manner. This is the basis of the various concentration tests for kidney function. The changes in the excretion of the inorganic constituents by the failing kidney have been emphasized less, but are more instructive here than the excretion of the organic constituents. In general the findings are that the inorganic constituents whose relative concentrations are greatest in normal urine, as compared with the blood, suffer the greatest reductions in the rates of excretion when the kidney is damaged. It is this and not the amount in the blood which appears to be the determining factor. For instance, Denis and Hobson's summary[41] of the results of their analysis of a fairly large number of cases of nephritis is: "The relative rarity of abnormally high figures for the sodium and chlorides in the blood would

therefore appear to show that these two ions are excreted by the kidney with great ease, whereas the retention of inorganic phosphates may amount to 400 per cent, and of sulfate to as much as 3000 per cent of the average normal value." They found that though the normal molar concentration of phosphates in the plasma is six times that of sulfate, in nephritis, the sulfate ion in the plasma became not only relatively, but, in some extreme cases, absolutely more concentrated than the phosphates. The normal ratios of the concentrations of the three anions, chloride, phosphate and sulfate in the urine and in the blood, is least for the chloride and greatest for the sulfate.**

It is unfortunate that the concentrations of the inorganic constituents of the urine in disease are not often given. In the experiments of Hendrix and Bodansky[42] on uranium nephritis in dogs, the data given show that the concentration of phosphates in the urine in all the animals that died fell to considerably below 0.1 per cent P_2O_5, and in the two animals that survived the concentrations, though lowered somewhat, returned to 0.1 per cent or more. Frandsen found in rabbits with uranium nephritis no such lowering of the concentration of chloride in the urine. Rather the reverse, the ability to retain chloride seemed to be impaired. This may have been due to the acidosis which is engendered in uranium poisoning, resulting in an increased excretion of chloride in compensation for the failure to excrete sulfate and phosphate and the inability to form ammonia.

To test more thoroughly this view of kidney function, it is planned to obtain more data on the concomitant variations in concentration of the constituents of the blood and the urine in health and disease.

It is possible that the efficiency of the kidney varies somewhat for each substance excreted, and further that the relationship between work performed and oxygen consumed is not linear. The available data does not seem to throw any light on this interesting problem which may be related to the problem of thresholds in renal excretion.

Summary.—1. The work of the kidney in the excretion of urine is analyzed by means of the second law of thermodynamics. It is shown that the production of either hypertonic or hypotonic urine entails work on the part of the kidney. The excretion of a urine which is the same as the plasma in all details incurs no work by the kidney.

2. The energy consumed by the kidney in man in the production of urine was found to be 6–11 kg. calories per gram of nitrogen excreted.

3. It is shown that the normal healthy kidney considered as a chemical machine possesses a great capacity for work; but performs its work with an "efficiency" probably not greater than 1–2 per cent.

4. The effect of disease is to reduce markedly the capacity of the kidney for work.

This conception of renal function is supported by the evidence in the literature of the constitution of the urine in health and disease; of the effects of drugs, metallic poisons, anaemia, cyanide and narcotics on the oxygen consumption of the kidney *in situ* and isolated, and on the constitution of the urine; and by the observations of the authors, confirming earlier observations, of the increase in metabolism in man during increased excretion of urea.

The authors wish to thank Professor R. C. Tolman for his interest in and assistance with this work.

The authors are indebted to the Pasadena General Hospital, and particularly to Miss Nena Lind and Dr. R. N. Crumrine for their generosity and hospitality in placing the metabolism laboratory of the hospital at their disposal.

* These definitions were suggested by Professor R. C. Tolman.

** From the observations of Davies, Haldane and Peskett[43] and of Harvard and Reay[44] it seems that the concentration of chloride in the urine in health varies between 3 times that in the plasma to only a trace. Harvard and Reay, by the ingestion of large quantities of water, reduced the concentration of phosphate in the urine to as low as $^1/_6$ that in the blood. This apparently is the lowest concentration of phosphate in the urine of healthy man on record. Even under these conditions of extreme dilution of the urine, urea and sulfate were found to be more concentrated in the urine than in the blood. Ignoring the question of threshold substances and the mechanism of renal excretion it may be stated that in the urine, chlorides normally are concentrated less and may be diluted more than phosphates. The relative concentration of sulfate is always greater than either of these.

BIBLIOGRAPHY

[1] Cushny, A. R., *The Secretion of Urine*, London, 1926, 37.

[2] Hill, A. V., quoted by Cushny, *Ibid.*, 39.

[3] Mathews, A. P., *Physiological Chemistry*, New York, 1927, 478.

[4] Hawk, P. B., and Bergeim, O., *Practical Physiological Chemistry*, Philadelphia, 1927, 594.

[5] Behre, J. A., and Benedict, S. R., *J. Biol. Chem.*, 1922, **52**, 11.

[6] Rehberg, P. B., *Biochem. J.*, 1926, **20**, 448.

[7] Sendroy, Jr., J., and Hastings, A. B., *J. Biol. Chem.*, 1926–27, **71**, 783.

[8] Henderson, L. J., *Blood*, New Haven, 1928, 43.

[9] Benedict, S. R., and Nash, T. P., *J. Biol. Chem.*, 1926, **69**, 381.

[10] Bliss, S., *J. Biol. Chem.*, 1926, **67**, 109.

[11] Bliss, S., *Ibid.*, 1928, **78**, viii.

[12] Bliss, S., *Ibid.*, 1929, **81**, 129.

[13] Lewis, G. N., and Randall, N., *Thermodynamics*, New York, 1923.

[14] Werner, E. A., *The Chemistry of Urea*, London, 1923.

[15] Barcroft, J., and Brodie, T. G., *J. Physiol.*, 1905, **32**, 18.

[16] Barcroft, J., and Brodie, T. G., *Ibid.*, 1905–06, **33**, 52.

[17] Barcroft, J., and Straub, H., *Ibid.*, 1910, **41**, 145.

[18] Knowlton, F. P., and Silverman, C. A., *Am. J. Physiol.*, 1918–19, **47**, 1.

[19] Bainbridge, F. A., and Evans, C. L., *J. Physiol.*, 1914, **48**, 278.

[20] Winfield, G., *Ibid.*, 1912–13, **45**, 182.

[21] Starling, E. H., and Verney, E. B., *Proc. Roy. Soc.*, 1925, **B97,** 321.
[22] Marshall, E. K., and Crane, M. M., *Am. J. Physiol.*, 1924, **70,** 465.
[23] Tamura, K., and Miwa, M., *Mittheil, med. Fakult. Kaiser, Univ. Tokyo*, 1920, **23,** 317, 349.
[24] Frandsen, J., *Scand. Archiv. Physiol.*, 1925, **46,** 193.
[25] Hayman, Jr., J. M., and Schmidt, C. F., *Am. J. Physiol.*, 1927–28, **83,** 502.
[26] Fee, A. R., *J. Physiol.*, 1929, **67,** 14.
[27] Fee, A. R., and Hemingway, A., *Ibid.*, 1928, **65,** 100.
[28] v. d. Heyde, H. C., *J. Biol. Chem.*, 1921, **46,** 421.
[29] Keilin, D., *Proc. Roy. Soc.*, 1929, **104B,** 206.
[30] Gray's *Anatomy*, London, 1923, 1192.
[31] Lusk, G., *The Elements of the Science of Nutrition*, Philadelphia, 1928, 65.
[32] Quoted from Lublin, A., *Zeit. Klin. Med.*, 1928–29, **109,** 371.
[33] Dubois, E. F., *Basal Metabolism in Health and Disease*, Philadelphia, 1927.
[34] Zuntz, N., *Med. Klin.*, 1910, 351.
[35] Lublin, A., *Zeit. Klin. Med.*, 1928–29, **109,** 371.
[36] Lusk, G., *J. Biol. Chem.*, 1912–13, **13,** 27.
[37] Grafe, E., *Deut. Archiv. Klin. Med.*, 1916, **118,** 1.
[38] Benedict, F. G., and Emmes, L. E., *Am. J. Physiol.*, 1912, **30,** 197.
[39] Tribe, E. M., Hopkins, F. G., and Barcroft, J., *J. Physiol.*, 1915–16, **50,** xi.
[40] Tribe, E. M., Harvey, W. H., and Barcroft, J., *Ibid.*, 1915–16, **50,** xii.
[41] Denis, W., and Hobson, S., *J. Biol. Chem.*, 1923, **55,** 183.
[42] Hendrix, B. M., and Bodansky, M., *Ibid.*, 1924, **60,** 657.
[43] Davies, A. W., Haldane, J. B. S., and Peskett, G. L., *J. Physiol.*, 1922, **56,** 269.
[44] Harvard, R. E., and Reay, G. A., *Biochem. J.*, 1926–27, **20,** 99.
[45] Mayrs, E. B., *J. Physiol.*, 1922, **56,** 58.
[46] Borsook, H., and Winegarden, H. M., *Proc. Nat. Acad. Sci.*, 1930, **16,** 559.
[47] Bayliss, W. M., *Principles of General Physiology*, London, 1927, 342.

THE THERMAL DECOMPOSITION OF NITROGEN PENTOXIDE AT LOW PRESSURES

BY J. H. HODGES AND E. F. LINHORST

DEPARTMENT OF CHEMISTRY, UNIVERSITY OF MICHIGAN

Communicated December 13, 1930

Introduction.—The rate of decomposition of nitrogen pentoxide at low pressures is of particular interest in connection with the theory of unimolecular reactions. Since Daniels and Johnston[1] first showed this reaction to be unimolecular, many investigations have demonstrated that the specific reaction rate is constant and reproducible, and that its unimolecular character is maintained over a wide range of conditions. The problem of the source of the energy of activation which a molecule must receive prior to its decomposition offers particular difficulties in the case of a unimolecular reaction. From the simplest point of view it would seem that this activating energy could not be received through the agency of impacts with

other molecules since the specific reaction rate is independent of the concentration and hence independent of the frequency of collision. But it was first pointed out by Lindemann[2] that it is possible to reconcile the fact that the specific reaction rate is apparently independent of the collision frequency with the idea that the energy of activation is acquired by impact with other molecules provided we assume that the average time elapsing between activation and decomposition is large compared with the average interval between two collisions. This idea has been elaborated by others (see references 4 and 5 in the paper by Hibben[5]). If we suppose that molecules may be deactivated as well as activated by impact with other molecules, an equilibrium condition will result in which a certain small fraction of all the molecules will be in the activated state and this fraction will be independent of the concentration, other conditions being equal. Under ordinary pressures, this constant proportion of activated molecules will be maintained so effectively by collisions that the rate of decomposition will depend entirely on the slow stage, the decomposition of the activated molecule, and the specific reaction rate will not be affected by comparatively large changes in the collision frequency as a result of changes in pressure. But it has been pointed out that a consequence of this theory would be a falling off in the specific reaction rate at very low pressures when the interval between collisions became of the order of magnitude of the time elapsing between activation and decomposition, and that when the pressure is so low that the rate of activation is distinctly less than the rate of decomposition of the activated molecule the reaction will become bimolecular since the activation process involves the collision of two molecules.

The results of previous work on the rate of decomposition of nitrogen pentoxide at low pressures are somewhat discordant. Hunt and Daniels[3] find the specific reaction rate essentially constant down to 0.01 mm., though the value at their lowest pressure seems to be slightly lower than at higher pressures. Hirst and Rideal[4] find the specific rate to increase at pressures below 0.26 mm., but no other work has indicated an increase at low pressures. Hibben[5] finds no falling off in the specific rate from 0.2 mm. down to 0.002 mm. Sprenger,[6] by a method similar in some respects to ours, comes to the remarkable conclusion that the reaction begins normally, slows down and stops before all the nitrogen pentoxide has decomposed. These conclusions seem rather improbable and are supported by no other work. Schumacher and Sprenger[7] report a decrease in the specific reaction rate below 0.06 mm. Ramsperger and Tolman[8] report careful measurements at pressures down to 0.001 mm. showing a decrease in the specific reaction rate at pressures below 0.05 mm., but giving no indication that the reaction changes its unimolecular character even at their lowest pressures. They also investigate the effect of increasing the glass surface exposed to the reacting system, and find a significant falling off in the specific reaction rate

in the course of a single run when the glass surface is increased five-fold.

Apparatus.—The reaction was carried out in a $1^1/_2$-liter pyrex flask having a ratio of surface-volume = 0.5 cm.$^{-1}$, and a volume, including the side tubes, of 1765 cc. The reaction vessel was cleaned with chromic acid before use. To the flask were attached a condensation tube, which could be immersed in liquid air, and the tube containing the manometer. We used the quartz fiber manometer of Haber and Kerschbaum as modified by Coolidge.[9] It consists in principle of a fine quartz fiber mounted rigidly at the top and oscillating freely at the bottom. The time during which the amplitude of oscillation decreases by one half depends only upon the pressure and the molecular weight of the gas filling the manometer, provided the pressure is not higher than a few hundredths of a millimeter. A given fiber can be calibrated in air against the readings of a McLeod gauge and from the calibration curve (or equation) the pressure of any gas of known composition may be determined by measuring with a stopwatch the time during which the amplitude decreases by one half. Between the range 0.001 mm. to 0.03 mm. we believe our pressure measurements are accurate to 2%. The reaction flask, manometer and condensation tube could be sealed off from the rest of the apparatus after filling, and placed in a thermostat to observe the course of the reaction.

For the operation of filling, the reaction system was sealed to the pump system, which included a side reservoir of nitrogen pentoxide, continually cooled by carbon dioxide snow, a liquid air trap, a McLeod gauge, and a two-stage mercury diffusion pump. At the outlet of the reservoir was placed an internal, greaseless, glass valve, carefully ground and actuated magnetically. Such a valve would not hold against large pressure differences, but gave every indication of being adequately tight against the very small pressure differences to which we subjected it. Another such valve was placed in the vacuum line between the reservoir outlet and the liquid air trap.

The nitrogen pentoxide was prepared by dehydrating pure nitric acid with phosphorus pentoxide and redistilling in a stream of ozonized oxygen into the reservoir tube, which was sealed permanently to the pump system and kept continually cool.

In addition to the reaction system of 1765 cc. capacity we used two other reaction vessels in which a much larger surface of glass was exposed to the reacting gas. These vessels consisted of a number of long side tubes sealed to a central tube and bent around it to make a compact unit. The advantage of this type of vessel over a flask packed with glass tubing was that temperature adjustments on removing from the thermostat would be much more rapid. One of these vessels had a volume of 215 cc. and a ratio of surface to volume = 5 cm^{-1}; the other, constructed of finer tubing, had a volume of 310 cc. and a ratio of surface to volume = 10 cm^{-1}. These vessels were cleaned with chromic acid as was the larger flask.

Experimental Procedure.—The reaction vessel was baked at 400°C. for at least two hours under a high vacuum. After the flask had cooled, the valves were set so that the reservoir communicated with the reaction system, the apparatus was filled with nitrogen pentoxide at a pressure of about 0.05 mm. and allowed to stand for an hour to remove any possible traces of oxidizable material or water vapor and to condition the glass surfaces. The flask was then immersed in cold water and darkened to prevent any initial decomposition; both valves were closed, shutting off the reaction system with an initial pressure of nitrogen pentoxide determined roughly by the temperature of the reservoir. This quantity of nitrogen pentoxide was then frozen out in the condensation tube by liquid air, the valve in the vacuum line was opened to complete the removal of uncondensed gases, and the reaction system sealed off.

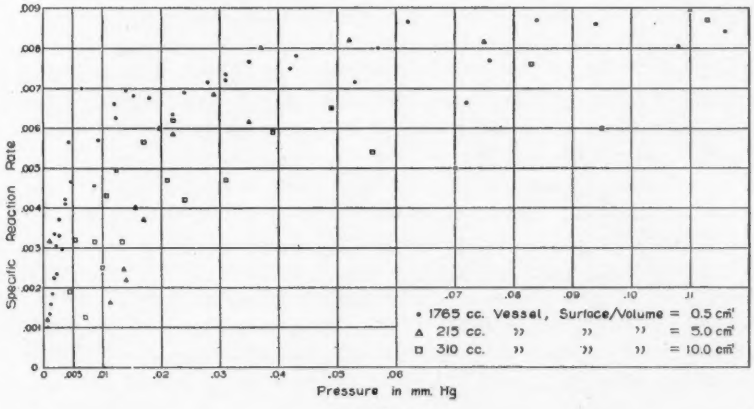


FIGURE 1

The flask was transferred to a water thermostat at 35.0°C. in a dark room, and the reaction allowed to proceed. At convenient intervals the flask was removed, the oxides of nitrogen frozen out rapidly with liquid air, and the oxygen pressure measured by the quartz manometer. At the end of a run complete decomposition was obtained by leaving the flask over night at 60°C.

Results.—Representative data obtained with the $1^1/_2$-liter flask are presented in tables 1 and 2; and all the results are expressed graphically in figure 1, where specific reaction rate for each interval is plotted against the average partial pressure of nitrogen pentoxide. The initial pressure of nitrogen pentoxide is calculated for each run from the final oxygen pressure. In table 1, which contains one representative run at the highest pressures we used and one run in our intermediate pressure range, the re-

action rate constant, K, is calculated for each time interval according to the usual formula for a unimolecular reaction, the time being expressed in minutes. In table 2, which contains the results of our three runs in the lowest pressure range, a unimolecular and a bimolecular constant are each calculated; for the latter expression, concentrations are expressed as partial pressure in millimeters of mercury.

TABLE 1

RUN NO.	TIME (MIN.)	O_2 PRESS. (MM. $\times 10^{-3}$)	N_2O_5 PRESS. (MM. $\times 10^{-3}$)	$K \times 10^3$ (UNIMOLEC.)
10	0	0.01	95.3	
10	30	11.0	73.4	8.7
10	70	21.8	51.9	8.65
10	120	29.9	35.7	7.5
10	180	36.1	23.2	7.15
10	∞	47.7	0	
6	0	0.05	16.8	
6	30	1.61	13.7	6.8
6	70	3.21	10.5	6.6
6	120	4.51	7.9	5.7
6	180	5.90	5.12	7.2
6	250	6.74	3.46	5.6
6	∞	8.47	0	

TABLE 2

RUN NO.	TIME (MIN.)	O_2 PRESS. (MM. $\times 10^{-3}$)	N_2O_5 PRESS. (MM. $\times 10^{-3}$)	$K \times 10^3$ (UNIMOLEC.)	K (BIMOLEC.)
18	0	0.02	5.37		
18	43	0.51	4.40	4.7	0.96
18	106	1.02	3.37	4.2	1.35
18	190	1.47	2.47	3.7	1.29
18	290	1.79	1.83	3.0	1.41
18	∞	2.685	0		Av. 1.25
8	0	0.01	4.15		
8	40	0.317	3.52	4.1	1.04
8	90	0.552	3.05	2.9	0.88
8	150	0.817	2.52	3.25	1.15
8	220	1.01	2.134	2.45	1.02
8	310	1.19	1.766	2.2	1.08
8	∞	2.086	0		Av. 1.04
9	0	0.00	1.88		
9	50	0.144	1.59	3.35	1.93
9	130	0.253	1.37	1.85	1.67
9	250	0.373	1.14	1.6	1.29
9	430	0.495	0.89	1.35	1.35
9	∞	0.940	0		Av. 1.56

Discussion.—For all our intervals above 0.06 mm. partial pressure of nitrogen pentoxide the average of six measurements gives $K = 0.00801$. Daniels and Johnston[1] find at higher pressures a value $K = 0.00808$ at 35°C. The decrease in the value of K begins at partial pressures of nitrogen pentoxide around 0.05 mm. to 0.06 mm. This is in agreement with

the observations of Schumacher and Sprenger[7] and of Ramsperger and Tolman.[8] The rate continues to fall more sharply until at pressures below 0.004 mm. the reaction becomes bimolecular as the values in the last column of table 2 indicate.

The results with the vessels of large surface, shown in figure 1, are less consistent than those with the $1^1/_2$-liter flask. They indicate a disturbing effect due perhaps to the glass surface; but it must be observed that the smaller volume of the two vessels of large surface may also have been a cause of results deviating from those obtained with the larger vessel. The only conclusion apparent from these results is that the reaction is slower over the entire range of pressures. We find no evidence of a more rapid reaction at higher pressures as did Ramsperger and Tolman[8] when they increased the glass surface.

The most significant point brought out in the above results is the apparent bimolecular character of the decomposition at pressures below 0.004 mm. Of the several papers reporting measurements in this pressure range, none gives any definite indication of a bimolecular reaction. The results of Schumacher and Sprenger,[7] while not so consistent as ours, are, however, of essentially the same nature. When their values for K are plotted against partial pressure of nitrogen pentoxide, the points lie in the same general region as those given in figure 1. We shall not attempt to discuss here the possible reasons for the discrepancies between our results at the lowest pressures and those in the papers just mentioned. Neither shall we discuss at present the important question of whether the apparent bimolecular character of the rate measurements in our apparatus is to be taken as meaning a true homogeneous reaction of the second order or whether possible surface effects may account for the results. The work is still in progress and such discussion can be undertaken with more profit when the results are complete.

Summary.—The rate of thermal decomposition of nitrogen pentoxide has been measured at 35°C. and at pressures between 0.12 mm. and 0.001 mm.

The reaction was found to be unimolecular above 0.06 mm. partial pressure of nitrogen pentoxide. Below 0.06 mm. the specific reaction rate drops steadily with decreasing pressure, and below 0.004 mm. the reaction appears to be bimolecular.

[1] Daniels and Johnston, *J. Am. Chem. Soc.*, **43**, 53–81 (1921).
[2] Lindemann, *Trans. Farad. Soc.*, **17**, 598–99 (1922).
[3] Hunt and Daniels, *J. Am. Chem. Soc.*, **47**, 1602–09 (1925).
[4] Hirst and Rideal, *Proc. Roy. Soc.* (*London*), **109A**, 526–40 (1925).
[5] Hibben, *J. Am. Chem. Soc.*, **50**, 940–50 (1928).
[6] Sprenger, *Z. physikal. Chem.*, **136**, 49–76 (1928).
[7] Schumacher and Sprenger, *Proc. Nat. Acad. Sci.*, **16**, 129–135 (1930).
[8] Ramsperger and Tolman, *Ibid.*, **16**, 6–13 (1930).
[9] Coolidge, *J. Am. Chem. Soc.*, **45**, 1637–43 (1923).

ON THE TRANSFER OF ENERGY BETWEEN ATOMS AT COLLISION

By Oscar Knefler Rice[1]

Chemical Laboratory, Harvard University

Communicated December 15, 1930

Recently Kallmann and London[2] have considered the question of energy exchange between atoms or molecules, and have concluded that it is possible to account theoretically for the fact that energy exchanges sometimes take place over distances much greater than are comparable with atomic or molecular radii. Their method is to consider that as one atom or molecule approaches another forces come into play which may be treated as perturbations, enabling one to calculate the rate at which energy exchanges take place as the molecules go by each other with different distances of approach, different velocities, etc. This amounts simply to calculating the rate at which shifts in the internal quantum states of the two molecules (considered together as one system) take place under the influence of the perturbation, which was considered to be due to the mutual action of the electric moments of the two atoms or molecules involved in the collision, as this gives the only possibility of getting a sufficiently large perturbation at large distances for appreciable energy exchange to take place.

In the method of Kallmann and London the relative translational energy of the colliding atoms or molecules is neglected completely. On this account it has been suggested by Frenkel[3] that Born's method of solving collision problems should be applied to this problem. But that such an application is not valid in this case because of the large perturbations involved has been pointed out by Kallmann and London.[4]

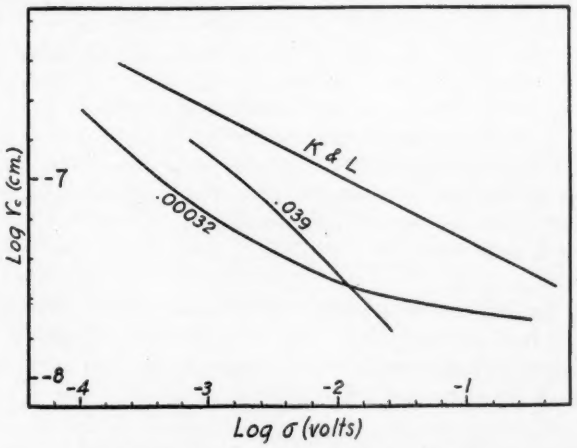
The question of the effect of the kinetic energy is nevertheless of fundamental importance, not only for this particular problem, but also for all the cases of energy exchange, which play such a rôle in chemical kinetics. The solution of the problem at hand may be considered the first step in the solution of many other problems. It therefore seems worth while to consider it anew, and in this note a modification of Born's method will be outlined, which takes into account the mutual kinetic energy of the particles and also meets the objections made by Kallmann and London to Frenkel's suggestion.[5] It thus far gives only an upper limit to the radius of action in any given case, but as this is already considerably smaller than that found by Kallmann and London and as the method should be of general use it seems worth while to present it now.

To do this we consider stationary states of a pair of atoms or molecules; the wave function ψ which describes such a state depends among other

variables on the coördinates giving the relative positions of the two atoms or molecules. In this note we shall consider only the case of two atoms, in which the inner electrons can be considered to be composed of smeared out shells, and only one electron, the "Leuchtelektron," is of importance. We let r', ϕ', and θ', and r'', ϕ'', and θ'' be the polar coördinates of the electrons on the first and second atoms respectively with respect to axes fixed in the respective nuclei; r, ϕ, and θ will be the polar coördinates of the line joining the two atoms with respect to a set of axes fixed in space, and parallel to the nuclear sets of axes. Then we may write (simply expressing the fact that the variables are separable)

$$\psi = Y'_{m',l'}(\phi', \theta')R'_{l',k'}(r')Y''_{m'',l''}(\phi'', \theta'')R''_{l'',k''}(r'')Y_{m,l}(\phi, \theta)R_{l,k}(r) \quad (1)$$

this being the zero approximation, the unperturbed solution. The sub-



scripts give the quantum numbers. l gives the angular momentum, m its component about the z-axis. It will be noted that all the parts of ψ have discrete sets of characteristic values except R, which due to the free motion of the two particles has a continuous set. It will be convenient to make this set discrete, also, by introducing a largest value r_1, of r.

The method consists in considering some given initial state, ψ_1, and finding what happens in the course of time under the influence of the perturbation due to the mutual action of the electric moments of the atoms. We can also find out what fraction of the pairs of molecules in the state ψ_1 will suffer collision in unit time, and thus we can find out the average effect of a collision. We can then average over all ψ_1's which correspond to a certain pair of internal states of the two molecules, and find out the rate at which they go to other internal states, in which an exchange of

internal energy has taken place. If the total amount of internal energy changes in this transition the difference must be taken up as change of translational energy. If this difference is zero we have Kallmann and London's case of exact resonance.

The formulation we have given is very convenient, for the following reason. R_k (we leave off the other subscripts) can be written as F_k/r where F_k obeys the differential equation

$$d^2F_k/dr^2 + (\eta^2E_k - l(l+1)r^{-2})F_k = 0 \qquad (2)$$

where $\eta^2 = 8\pi^2M/h^2$; where M is the reduced mass, E_k is the translational energy. Now it is seen that F_k decreases rapidly as we go to values of r less than $r_0 = [l(l+1)/\eta^2E_k]^{1/2}$ and this distance may be taken as the distance of closest approach of two molecules in collision, and it is determined by l and E_k (i. e., by their *initial* values, as shown hereafter by subscript 1).

We shall now consider the rate of transfer from any state ψ_1 to another definite state ψ_2 of the same total energy. We denote the perturbation as a function of the coördinates as V and the perturbation matrix component as v_{12}. The ψ's are normalized so that $\int \psi^2 d\tau = 1$ in which r goes from 0 to r_1, all the other coördinates being taken over the usual limits, $d\tau$ being the volume element. Then if v_{12} is not too large, the rate of transfer by collisions may be shown to be[6]

$$\gamma_{12} = 4\pi^2v_{12}^2/h\epsilon_2, \qquad (3)$$

where ϵ_2 is the distance between two "continuous" states which go with the final internal pair of states, being given by $hE_{k_2}^{1/2}/r_1(2M)^{1/2}$. Now it can be shown that the number of collisions per unit time is ϵ_1/h, hence the chance of transition per collision is

$$4\pi^2 v_{12}^2/\epsilon_1\epsilon_2. \qquad (4)$$

If this becomes of the order of magnitude of 1 this, of course, no longer holds. The transition probability per collision can never be greater than 1. If we find the value of r_0 for which the expression (4) becomes equal to 1, this will give us roughly a maximum value for the effective radius of the given transition.[7]

We do this with a given initial internal state and a given E_{k_1} (which we will take as the average relative energy of transition, and simply use the result instead of averaging over E_{k_1}). Then the value of r_0 is determined by l_1, which likewise determines F_{k_1} by Eq. (2), which determines v_{12}. In order to find the critical value of r_0 properly one should find the average value of v_{12}^2 (averaged over all m_1) for any given l_1, and one should also add together all values of v_{12}^2 if more than one final state is possible with the initial state given. When we speak of v_{12}^2 hereafter we will mean this sum.

The method of procedure is now definitely outlined. All that remains is to determine v_{12}. As V is the energy due to the mutual action of the electric moments due to the relative positions of the leucht-electrons and the smeared-out central parts of the atoms it is given by (the atoms being far apart compared to their sizes)

$$V = 2\xi'\xi''\,[\cos\theta'\cos\theta''\cos^2\theta + \sin\theta'\cos\theta''\sin\theta\cos\theta\cos(\phi' - \phi) + \cos\theta'\sin\theta''\sin\theta\cos\theta\cos(\phi'' - \phi) + \sin\theta'\sin\theta''\sin^2\theta\cos(\phi' - \phi)\cos(\phi'' - \phi)]r^{-3} \quad (5)$$

where ξ' and ξ'' are the absolute magnitudes of the electric moments of the respective atoms and are functions of r' and r'' only. The integration with respect to r' and r'' will introduce in v_{12}^2 a factor e^4a^4 where a is of the order of 10^{-8} cm. and e is the charge on the electron. The integration with respect to the angles brings out certain selection rules for the corresponding quantum numbers. We have $l_2' = l_1' \pm 1$, $l_2'' = l_1'' \pm 1$, $l_2 = l_1$ or $l_1 \pm 2$, $m_2' = m_1'$ or $m_1' \pm 1$, $m_2'' = m_1''$ or $m_1'' \pm 1$, $m_2 = m_1$ or $m_1 \pm 1$ or $m_1 \pm 2$. Further we must always have $m_2' + m_2'' + m_2 = m_1' + m_1'' + m_1$ giving conservation of angular momentum about the z-axis. The method used in this integration is that described by Sommerfeld.[8] The integration with respect to r can be performed as the r-part of the eigenfunction can be expressed as a Bessel function, the integral then being of a standard form.[9] The value of this part of the integral depends on whether l changes by 0 or ± 2.

In setting forth the results we may consider three cases:

Case 1: The case of exact resonance.

Case 2: $E_{k_1}/E_{k_2} < 1 - 2(r_0\eta E_{k_1})^{-1/2}$. In this case only transitions in which $l_2 = l_1 + 2$ contribute appreciably to the result.

Case 3: $E_{k_2}/E_{k_1} < 1 - 2(r_0\eta E_{k_1})^{-1/2}$. In this case only transitions in which $l_2 = l_1 - 2$ contribute appreciably.

For practical purposes, these three cases are sufficient, and we can give very approximately the results obtained by setting expression (4) equal to 1. r_c is the value of r_0 thus obtained.

Case 1: $r_c = 0.36\, ae\eta^{1/2}E_{k_1}{}^{-1/4}$

Case 2: $1 = 0.0024\, a^4e^4\eta^5E_{k_1}{}^{1/2}(E_{k_1}/E_{k_2})^{r_c\eta E_{k_1}{}^{1/2}-1}(1 - E_{k_1}/E_{k_2})^3 r_c{}^{-1}$

Case 3: $1 = 0.0024\, a^4e^4\eta^6E_{k_1}(E_{k_2}/E_{k_1})^{r_c\eta E_{k_1}{}^{1/2}-1}(1 - E_{k_2}/E_{k_1})^3(r_c\eta E_{k_1}{}^{1/2} - 2)^{-1}$

In Case 1 the expression is essentially the same as that of Kallmann and London, except that our cross section (not radius) is smaller by a factor of about 9. It is difficult to gauge the importance of such a factor. It does not seem to me that the mechanism considered can account for

the results of Hanle[10] and Datta[10] on the depolarization of resonance radiation in any event,[11] as they found much larger cross sections. For the other cases I do not think that a factor of the size noted makes much difference, as far as the meager present data for the case of exact resonance goes. The difference is due to the integration with respect to the angles, which Kallmann and London left out.

The results in Cases 2 and 3 have no outward resemblance to those of Kallmann and London and in every case show a considerably smaller r_c for given $E_{k_2} - E_{k_1}$. In order to show how r_c varies with $\sigma = E_{k_2} - E_{k_1}$, I have made two graphs for Case 2, taking $a = 10^{-8}$ cm., $E_{k_1} = 0.039$ and 0.00032 volt electrons respectively, and $M = 20$ atomic weight units. For purposes of comparison I have also included the curve for the case $a = 10^{-8}$ cm., $E_{k_1} = 0.039$ volts and $M = 20$ which one obtains from Kallmann and London's formula.[12] I have carried the curve to lower values of r_c than those for which the above calculation really gives a good approximation.

One notes that the curves for 0.039 and 0.00032 volts differ somewhat, but not sufficiently but that one appears to be justified in using, at any temperature, simply the curve in which E_{k_1} is given its average value for that temperature.

Unfortunately the experimental data which is at hand is mostly for complicated systems, and does not lend itself readily to comparison with the above. There are many instances, however, in which the cross sections are surprisingly large, and at present I cannot see how any slight modification or adaptation of the present theory could account for them. It may be that some rather fundamental change will be needed to explain them. On the other hand, it seems that it should be possible to find cases where the above theory will apply.

As I have stated the above results give upper limits to the value of r_c for a given σ. Preliminary calculations,[7] for the case where the perturbations are large, indicate that r_c is probably even considerably less, and that exceedingly good resonance is necessary for any energy exchange to take place.

A detailed account of the above research will be published later elsewhere. It will be seen that it ought to be possible to adopt the method to other cases and an attempt will be made to do this.

In conclusion I wish to thank Professor Heisenberg for helpful discussion, and for courtesies extended me during my stay in Leipzig. I have also had the privilege of discussing part of the above work with Dr. Kallmann, who informs me that he has also independently worked out a more exact treatment than that previously given, which is in some respects similar to the above, though not exactly the same, it giving, I understand, Kallmann and London's old result as a first approximation.

I have also had several stimulating discussions with Dr. Clarence Zener.

[1] This work was begun while the writer was a National Research Fellow at the Institute for Theoretical Physics of the University of Leipzig.

[2] Kallmann and London, *Z. physik. Chem.*, **2B**, 207 (1929).

[3] Frenkel, *Z. Physik*, **58**, 794 (1929).

[4] Kallmann and London, *Ibid.*, **60**, 417 (1930).

[5] A perfectly general method for treating a similar problem has been given by Holtsmark, *Z. Physik*, **52**, 485 (1928). This method, however, cannot be used here for actual calculations.

[6] Dirac, *Proc. Roy. Soc.*, **114A**, 256 (1927).

[7] If the transition is not one of exact resonance there will be certain matrix components (corresponding to transitions in which no or little change of kinetic energy takes place and which, therefore, cannot occur with conservation of energy) very much larger than v_{12}. It does not seem likely this could make the probability of transition to the state 2 itself any greater, as this will not combine with any of the states (not conserving energy) which one would expect to be temporarily excited while the two atoms are close together. Indeed, a preliminary calculation indicates, rather, that the probability of transition would be lessened by such disturbing influences.

[8] Sommerfeld, *Atombau und Spektrallinien, Wellenmechanischer Ergänzungsband*, Vieweg und Sohn, Braunschweig, 1929, p. 62, ff.

[9] Watson, *Theory of Bessel Functions*, Cambridge, 1922, p. 401.

[10] Datta, *Z. Physik*, **37**, 625 (1926). Hanle, *Ibid.*, **41**, 164 (1927).

[11] As v_{12}^2 falls off so rapidly with the distance, I see no hope in the cumulative effect of the small perturbations due to atoms at a great distance from each other. Compare Kallmann and London, loc. cit. (*Z. physik. Chem.*), 233, ff.

[12] Kallmann and London, loc. cit. (*Ibid.*) Eq. (26).

GROUPS WHICH ADMIT FIVE-EIGHTHS AUTOMORPHISMS

BY G. A. MILLER

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS

Communicated December 3, 1930

In two previous articles[1] we determined fundamental properties of all the groups which admit three-fourths automorphisms as well as of all those which admit two-thirds automorphisms. In the present article similar properties are determined as regards the groups which admit five-eighths automorphisms. It is also proved here that there are no groups which admit automorphisms between three-fourths and two-thirds, or between two-thirds and five-eighths. The former of these two facts may be established as follows: If a group G admits an l automorphism where l is less than three-fourths and more than two-thirds, then G must be non-abelian and it must involve two non-commutative operators s_α and s_β which correspond to their inverses in such an automorphism.

Let H_α and H_β be the subgroups composed of all the operators of G

which are commutative with s_α and s_β, respectively. Since not both of these subgroups can be abelian and also of index 2 we may assume that H_β does not have both of these properties. It is easy to prove that H_β could not be non-abelian. If it were non-abelian then at least one-fourth of its operators could not correspond to their inverses under the automorphism in question. Hence the number of products of s_β into operators of G which would not correspond to their inverses in this automorphism would be at least equal to the number l of operators which correspond to their inverses but are not found in $H_\beta + h_\beta/4 = l - h_{\beta/2}$, h_β being the order of H_β. This is clearly impossible when $l > 5g/8$. Hence we have proved that H_β cannot be non-abelian whenever the automorphism of G exceeds 5/8. It will be seen later that H_β can be non-abelian when the automorphism is 5/8.

When H_β is abelian and of index 3 the automorphism can obviously not exceed 2/3 and when the index exceeds 3 the automorphism must be less. It has therefore been proved that the automorphism of G could not exceed 2/3 and be less than 3/4. To prove that it could also not exceed 5/8 and be less than 2/3 it remains only to consider the case when H_β is abelian. If H_β were invariant and of index 3 then s_α could not be commutative with one-half of the operators of H_β since it could not transform H_β into an automorphism of order 2. If s_α were commutative with less than one-half of the operators of H_β then at least 4/9 of the operators of G would not correspond to their inverses. Hence H_β could not be both abelian and invariant if it is of index 3. When it is non-invariant and abelian G admits a two-thirds automorphism, and if its index exceeds 3 the automorphism could clearly not exceed five-eighths. It has therefore been proved that *there exists no group which admits an automorphism which is either less than three-fourths and greater than two-thirds, or less than two thirds and greater than five-eighths.*

In the articles to which we referred it was proved that if a group admits either a three-fourths automorphism or a two-thirds automorphism it must involve an abelian subgroup of index 2. It is easy to prove as follows that this is not necessarily true of groups which admit five-eighths automorphisms. Let H be any group which admits three-fourths automorphisms and has an order which is divisible by 16, and suppose that this H is composed of all the operators of G which are commutative with s_α while s_β transforms half of the operators of the central of H into themselves multiplied by the commutator of H and is commutative with half of the operators of H. Moreover, assume that s_β^2 is in the central of G. It then results directly that a three-fourths automorphism of H gives rise to a five-eighths automorphism of G and that the commutator subgroup of G is of order 2 while its group of inner isomorphisms is the abelian group of order 16 and of type (1, 1, 1, 1). Hence it results that *every group*

which admits three-fourths automorphisms and has an order which is a multiple of 16 *can be extended to a group of twice its order which admits five-eighths automorphisms, has a commutator subgroup of order* 2, *and does not involve an abelian subgroup of index* 2.

When H_α is non-abelian in a five-eighths automorphism of G it is obvious that its index must be 2 and that s_α must transform it in the manner noted in the preceding paragraph in order to obtain a group which does not involve an abelian subgroup of index 2, since one-half of the products obtained by multiplying the operators of H_α by s_β must correspond to their inverses in such an automorphism. We proceed to prove that if a group does not involve an abelian subgroup of index 2 but admits at least one five-eighths automorphism it must belong to the category described in the preceding paragraph. It is easy to show that if this were not the case there would be a group such that in one of its five-eighths automorphisms all the subgroups composed of the operators which are commutative separately with non-invariant operators corresponding to their inverses would be abelian and of index 4. Hence all the products of such a non-invariant operator into the operators which do not correspond to their inverses in this automorphism would correspond to their inverses therein.

Hence it results that the three-eighths of the operators of G which do not correspond to their inverses in the five-eighths automorphism in question generate a group all of whose operators are invariant under the groups generated by the products formed by multiplying one of the non-invariant operators which correspond to their inverses into each of the others which have these properties. As both of these groups would have an index under G which does not exceed 2 they would have in common a subgroup whose index could not exceed 4. It therefore results that G involves an abelian subgroup of index 2 and hence it has been proved that *if a group admits a five-eighths automorphism but does not involve an abelian subgroup of index* 2 *it must involve a subgroup of this index which admits three-fourths automorphisms, and it must belong to the category determined above.* It remains therefore only to determine those groups which admit a five-eighths automorphism and contain an abelian subgroup of index 2.

Let G be a group which admits a five-eighths automorphism and involves an abelian subgroup H of index 2. It will first be noted that it is not possible to make exactly one-half of the operators of H correspond to their inverses in such an automorphism. If exactly one-half of the operators of H correspond to their inverses in an automorphism of H the rest of its operators must correspond to their inverses multiplied by a common operator of order 2. If s_α is an operator of G which is not contained in H but corresponds to its inverse in a five-eighths automorphism of G it results that three-fourths of the products obtained by multiplying s_α into H must correspond to their inverses. This is not

possible since s_α would have to transform every operator of H either into itself or into itself multiplied by the same operator of order 2. We have therefore proved the following theorem: *If a group which involves an abelian subgroup of index 2 admits a five-eighths automorphism, then either all of the operators of this subgroup or exactly one-fourth of them correspond to their inverses in this automorphism.*

When each of the operators of the abelian subgroup H of index 2 under G corresponds to its inverse in a five-eighths automorphism of G, then one of the remaining operators s of G which corresponds to its inverse in this automorphism must be commutative with all the operators of a subgroup K of index 4 contained in H. When the quotient group H/K is non-cyclic K must involve the non-cyclic group of order 4, and when K involves such a non-cyclic subgroup and the quotient group H/K is non-cyclic such an automorphism can be established. Hence a necessary and sufficient condition that G admits such a five-eighths automorphism is that the invariants of H which are powers of 2 satisfy one of the following three conditions: At least two of these invariants exceed 2, one of them exceeds 2 and at least two others are equal to 2, or at least four of them are equal to 2. When H/K is the cyclic group of order 4 then the corresponding multiplying subgroup of order 4 must be cyclic and its operators of order 4 must correspond to their inverses in this automorphism of order 2. This implies that the invariants of H which are powers of 2 satisfy one of the following three conditions: At least one of them is 8, at least one of them is greater than 8 and at least one is less than 8, or at least two of them are equal to 4.

From the preceding paragraph it results that the invariants of H which are not powers of 2 do not affect the usefulness of H for the purpose of establishing a five-eighths automorphism of G. It also results that there are at least four such five-eighths automorphisms whenever there is one since any of the co-sets of G with respect to K which is not in H may be composed of operators which correspond to their inverses. All the groups which admit such a five-eighths automorphism admit also a five-eighths automorphism in which just one-fourth of the operators of H correspond to their inverses and each of the operators of G which is not also in H satisfies the same condition. This automorphism is obviously characteristic. Hence we have established the following theorem: *If a group which involves an abelian subgroup of index 2 admits one five-eighths automorphism it admits exactly five such automorphisms and its group of inner isomorphisms is of order 8.* This group of order 8 is either the octic group or the abelian group of type (1,1,1).

The dihedral group of order $4k$ clearly admits a $(k + 1)/2k$ automorphism in which all the operators correspond to themselves as well as k such automorphisms in which all the operators of the cyclic subgroup of index 2

correspond to their inverses. Hence such a group admits $k + 1$ $(k + 1)/2k$ automorphisms, where k is an arbitrary natural number. These automorphisms are distinct whenever $k > 1$. For $k = 2$, 3 and 4 there result groups which admit three-fourths, two-thirds and five-eighths automorphisms respectively. The group of inner isomorphisms of all of these groups is the dihedral group of order $2k$. When $k = 2$ or 3 it has been proved that all the groups which admit $(k + 1)/2k$ automorphisms have such a group of inner isomorphisms but this is not necessarily true when $k = 4$, as was proved above. In fact, in this case the group of inner isomorphisms may also be abelian and of order 8 and of type (1, 1, 1) as well as of order 16 and of type (1, 1, 1, 1). Hence the groups which admit five-eighths automorphisms are composed of various categories of elementary groups which have this useful common property.

[1] These PROCEEDINGS, **15** (1929), 369; **16** (1930), 86.

MULTIVECTORIAL CURVATURE

BY M. S. KNEBELMAN

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY

Communicated December 12, 1930

Throughout what follows I consider an n-dimensional space V_n with a Riemannian metric which I assume to be positive-definite—a restriction which is imposed for simplicity's sake and which may be removed by a slight change in the definition of length.

In a previous paper[1] I defined an r-cell determined by $r(\leq n)$ linearly independent vectors λ^i_α, $(\alpha = 1, \ldots, r)$ indicating the vector and $(i = 1, \ldots, n)$ indicating the component. It is the object of this note to define the curvature of an r-cell or of an r-plane and to obtain some of its properties.

Let the space V_n be referred to an arbitrary coordinate system x and let y be a Riemannian coördinate system with origin at the point x_0. Then the equations of any geodesic through x_0 are $y^i = \lambda^i(x_0)\, t$, λ^i determining the direction of the curve.

Consider the set of all geodesics through x_0 whose directions are linearly dependent on the r independent directions $\lambda^i(x_0)$. Any r independent geodesics in this set will have the equations

$$y^i = c^\beta_\alpha \lambda^i_\beta t^\alpha = \lambda^i_\beta(x_0)\, \tau^\beta \qquad (1)$$

the determinant $|c^\beta_\alpha|$ being different from zero. The above equations define a subspace V_r of V_n whose induced metric,[2] $b_{\alpha\beta}$ is given by

$$b_{\alpha\beta} = g_{ij} \lambda^i_\alpha \lambda^j_\beta \qquad (2)$$

g_{ij} being the components of the fundamental tensor of V_n and the summation convention being used for all repeated indices as well as for subscripts of indices. If we denote the components of the Riemann-Christoffel tensor of V_n by R_{ijkl} and those for the subspace V_r by $B_{\alpha\beta\gamma\delta}$, then it can be shown easily that

$$B_{\alpha\beta\gamma\delta} = R_{ijkl}\,\lambda^i_\alpha\lambda^j_\beta\lambda^k_\gamma\lambda^l_\delta. \tag{3}$$

The space V_n being non-singular at x_0 and the vectors being independent, $b \equiv |b_{\alpha\beta}| > 0$; in fact, b is the square of the content[1] of the r-cell. If $b^{\alpha\beta}$ denotes the normalized cofactor of $b_{\alpha\beta}$ in b, then $b^{\alpha\beta}.b^{\gamma\delta}.B_{\alpha\gamma\beta\delta}$ is the scalar curvature of V_r. We now define *the r-vectorial curvature of the space V_n for the orientation λ^i_α, $\alpha = 1, \ldots, r (1 < r \leq n)$ to be one-half the scalar curvature of the r-cell determined by λ^i_α.*

The factor one-half in the above definition is not essential and is used merely to make this definition agree with that of Riemannian curvature ($r = 2$). We shall denote this curvature by $K_{(r)}$ or simply by K if there is no chance for confusion. From (3) it follows that K is a scalar invariant and it can be shown to remain unchanged if instead of the orientation λ^i_α we use a linearly dependent one. Hence if we define an r-plane to be the totality of r-cells whose orientations are dependent on λ^i_α, we may properly speak of the curvature of an r-plane; in this r-plane we may choose the set λ^i_α to be mutually orthogonal unit vectors; that is, $g_{ij}\lambda^i_\alpha\lambda^j_\beta = \delta_{\alpha\beta}$. With this choice we have

$$K_{(r)} = {}^1/_2\,\delta^{\alpha_1\beta_1}_{\delta^{\alpha_2\beta_2}}\,B_{\alpha_1\alpha_2\beta_1\beta_2} = \sum_{\alpha_1 < \alpha_2}^{r} B_{\alpha_1\alpha_2\alpha_1\alpha_2}$$

If we denote by ${}_\mu K_{(r-1)}$ the $(r-1)$ dimensional curvature of the face of the r-cell obtained by omitting the vector λ^i_μ, we get

$$K_{(r)} = \frac{1}{r-2}\sum_{\mu=1}^{r} {}_\mu K_{(r-1)}. \tag{4}$$

We can therefore state the theorem due to Cartan.[3]

I. *The Riemannian curvature of an r-cell $(2 < r \leq n)$ is equal to $1/(r-2)$ times the sum of the curvatures of its $(r-1)$ dimensional faces.*

Using (4) as a recursion formula we have

$$K_{(r)} = \frac{1}{r-2}\cdot\frac{2}{r-3}\sum_{\mu_1<\mu_2}^{\binom{r}{r-2}} {}_{\mu_1\mu_2}K_{(r-2)} = \ldots = \sum^{\binom{r}{2}} K_{(2)}$$

From this it follows that

II. *If a space V_n has constant two-dimensional curvature its curvature of every dimension is constant.*

The inverse of this theorem is also true with two possible exceptions.

To prove this we shall state and prove a theorem which is a generalization of a theorem of F. Schur.[4]

III. *If the r-dimensional curvature of a space V_n, $1 < r \leq n-1$, is independent of orientation, it does not vary from point to point.*

To this end we may write $K = {}^1/_4\, b^{\alpha_1\alpha_2;\,\beta_1\beta_2}\, B^{\alpha_1\alpha_2;\,\beta_1\beta_2}$, where $b^{\alpha_1\alpha_2;\,\beta_1\beta_2}$ is a two-rowed minor of $|b^{\alpha\beta}|$ and by a theorem on compound determinants[5] we get

$$K = \frac{\epsilon^{\alpha_1\ldots\alpha_r}\,\epsilon^{\beta_1\ldots\beta_r}\,B^{\alpha_1\alpha_2\beta_1\beta_2}\,b_{\alpha_3\beta_3}\ldots b_{\alpha_r\beta_r}}{4\cdot b\cdot(r-2)!} \tag{5}$$

Upon replacing $b_{\alpha\beta}$ and $B_{\alpha\beta\gamma\delta}$ by their values from (2) and (3) this becomes

$$K = \frac{\epsilon^{\alpha_1\ldots\alpha_r}\,\epsilon^{\beta_1\ldots\beta_r}\,R_{i_1i_2j_1j_2}\,g_{i_3j_3}\ldots g_{i_rj_r}\prod_{\mu=1}^{r}\lambda^{i_\mu}_{\alpha_\mu}\lambda^{j_\mu}_{\beta_\mu}}{4\,(r-2)!\,g_{k_1\ldots k_r;\,l_1\ldots l_r}\prod_{\mu=1}^{r}\lambda^{k_\mu}_{\mu}\lambda^{l_\mu}_{\beta_\mu}} \tag{6}$$

The numerator of the right hand side of (6) may be transformed by the theory of determinants[6] into

$$\epsilon_{\alpha_1\ldots\alpha_r}\,\epsilon_{\beta_1\ldots\beta_r}\,R_{i_1i_2j_1j_2}\,g_{i_3j_3}\,g_{i_4j_4}\ldots g_{i_rj_r}\prod_{\mu=1}^{r}\delta^{i_\mu\alpha_\mu}_{a_\mu}\delta^{j_\mu\beta_\mu}_{b_\mu}\prod_{\mu=1}^{r}\lambda^{a_\mu}_{\mu}\lambda^{b_\mu}_{\mu}$$

The tensor $\dfrac{\epsilon_{\alpha_1\ldots\alpha_r}\,\epsilon_{\beta_1\ldots\beta_r}\,R_{i_1i_2j_1j_2}\,g_{i_3j_3}\ldots g_{i_rj_r}\prod_{\mu=1}^{r}\delta^{i_\mu\alpha_\mu}_{a_\mu}\delta^{j_\mu\beta_\mu}_{b_\mu}}{4\,(r-2)!}$ we denote by $R_{a_1\ldots a_r;\,b_1\ldots b_r}$.

If we raise the second set of indices by means of g^{bc} we get

$$R^{c_1\ldots c_r}_{a_1\ldots a_r} = \delta^{i_1\ldots i_r}_{a_1\ldots a_r}\,g^{j_1\ldots j_r;\;c_1\ldots c_r}\,R_{i_1i_2j_1j_2}\,g_{i_3j_3}\ldots g_{i_rj_r}/\,4(r-2)!,$$

which may be further transformed into

$$R^{c_1\ldots c_r}_{a_1\ldots a_r} = \frac{\delta^{c_1\ldots c_r\;rsk_3\ldots k_r}_{pqk_3\ldots k_r}\,\delta^{pq}_{a_1\ldots a_r}\,R^{pq}_{rs}}{4\,(r-2)\,!}. \tag{7}$$

From (7) it is evident that $R^{a_1\ldots a_r}_{c_1\ldots c_r}$ is skew-symmetric in both sets of indices and that it vanishes if more than two contravariant indices do not occur among the covariant ones.

Equation (6) may now be written as

$$Kg_{i_1\ldots i_r;\,j_1\ldots j_r}\prod_{\mu=1}^{r}\lambda^{i_\mu}_{\mu}\lambda^{j_\mu}_{\mu} = R^{k_1\ldots k_r}_{j_1\ldots j_r}\,g_{i_1k_1}\ldots g_{i_rk_r}\prod_{\mu=1}^{r}\lambda^{i_\mu}_{\mu}\lambda^{j_\mu}_{\mu} \tag{8}$$

Making the assumption of our theorem, viz., that K is independent of λ^i_α, we get by differentiating (8) with respect to $\lambda^{a_1}_1\lambda^{b_1}_1\ldots\lambda^{a_r}_r\lambda^{b_r}_r$ successively,

$$Kg_{i_1 \ldots i_r; j_1 \ldots j_r} \prod_{\mu=1}^{r} (\delta^{i_\mu}_{a_\mu} \delta^{j_\mu}_{b_\mu} + \delta^{i_\mu}_{b_\mu} \delta^{j_\mu}_{a_\mu}) = R^{k_1 \ldots k_r}_{j_1 \ldots j_r} \prod_{\alpha=1}^{r} g_{i_\alpha k_\alpha} \prod_{\mu=1}^{r} (\delta^{i_\mu}_{a_\mu} \delta^{j_\mu}_{b_\mu} + \delta^{i_\mu}_{b_\mu} \delta^{j_\mu}_{a_\mu}). \quad (9)$$

Multiplying each side of (9) by $\prod_{\alpha=3}^{r} g^{a_\alpha b_\alpha}$ and contracting the a's and the b's we get after considerable reduction

$$\frac{(n-2)!}{(n-r)!} Kg_{i_1 i_2;\ j_1 \ldots\ldots j_2} (\delta^{i_1}_{a_1} \delta^{j_1}_{b_1} + \delta^{i_1}_{b_1} \delta^{j_1}_{a_1})(\delta^{i_2}_{a_2} \delta^{j_2}_{b_2} + \delta^{i_2}_{b_2} \delta^{j_2}_{a_2}) = g_{i_1 h_1} g_{i_2 h_2} R^{h_1 h_2 j_3 \ldots j_r}_{j_1 j_2 j_3 \ldots j_r} (\delta^{i_1}_{a_1} \delta^{j_1}_{b_1} + \delta^{i_1}_{b_1} \delta^{j_1}_{a_1})(\delta^{i_2}_{a_2} \delta^{j_2}_{b_2} + \delta^{i_2}_{b_2} \delta^{j_2}_{a_2}). \quad (10)$$

From the symmetry properties of R mentioned above we can easily show that $R^{h_1 h_2 j_3 \ldots j_r}_{j_1 j_2 j_3 \ldots j_r}$ must be of the form

$$A_1 R^{h_1 h_2}_{j_1 j_2} + A_2 R\, \delta^{h_1 h_2}_{j_1 j_2} + A_3 (\delta^{h_1}_{j_1} R^{h_2}_{j_2} + \delta^{h_2}_{j_2} R^{h_1}_{j_1} - \delta^{h_1}_{j_2} R^{h_2}_{j_1} - \delta^{h_2}_{j_1} R^{h_1}_{j_2})$$

and to compute the values of these three coefficients we divide the terms of $R^{h_1 \ldots h_r}_{j_1 \ldots j_r}$ into groups which upon contraction give the tensors of the required type. Making use of the relations between the generalized Kronecker deltas[6] we find

$$A_1 = (r-2)!\binom{n-4}{r-2};\ A_2 = \frac{(r-2)!}{2}\binom{n-4}{r-4};\ A_3 = (r-2)!\binom{n-4}{r-3}. \quad (11)$$

Multiplying (10) by $g^{a_2 b_2}$ and contracting we find

$$R^i_j = 2\frac{n-1}{r(r-1)} K\delta^i_j \text{ and hence } R = 2\frac{n(n-1)}{r(r-1)} K. \quad (12)$$

Equations (10) may now be reduced to

$$A_1 (R_{a_1 a_2 b_1 b_2} + R_{a_1 b_2 b_1 a_2}) = \frac{2(r-2)!\binom{n-4}{r-2}}{r(r-1)} K (g_{a_1 a_2;\ b_1 b_2} + g_{a_1 b_2;\ b_1 a_2}). \quad (13)$$

Now if $R_{a_1 a_2 b_1 b_2} + R_{a_1 b_2 b_1 a_2} = S_{a_1 a_2 b_1 b_2}$ it may be shown by means of the identities existing between the components of the Riemann-Christoffel tensor that $R_{a_1 a_2 b_1 b_2} = {}^1/_3(2S_{a_1 a_2 b_1 b_2} + S_{a_1 b_1 a_2 b_2})$. Applying this to (13) we get—provided $A_1 \neq 0$—

$$R_{a_1 a_2 b_1 b_2} = \frac{2K_{(r)}}{r(r-1)} g_{a_1 a_2;\ b_1 b_2}.$$

It therefore follows from the theorem of Schur and by II that $K_{(r)}$ is independent of x.

It only remains to consider what may be said about the space V_n when $A_1 = 0$. Now $\binom{n-4}{r-2} = 0$ has only two solutions $r = n$ or $r = n-1$. If $r = n$ let μ^α_i be the normalized cofactor of λ^i_α in $|\lambda^i_\alpha|$; then $b^{\alpha\beta} = g^{ij} \mu^\alpha_i \mu^\beta_j$

$$K_{(n)} = {}^1/_2\, g^{ij}\, \mu_i^\alpha\, \mu_j^\beta\, g^{kl}\, \mu_k^\gamma\, \mu_l^\delta\, R_{pqrs}\, \lambda_\alpha^p\, \lambda_\gamma^q\, \lambda_\beta^r\, \lambda_\delta^s = {}^1/_2\, R$$

which is obviously independent of orientation so that the theorem need not be true in this case. If $r = n - 1$ let λ_n^i be any vector which is not linearly dependent on the set λ_α^i. Then (8) may be written in this case as

$$g \cdot g^{ij}\, \mu_i^n\, \mu_j^n\, K_{(n-1)} = R_{i_1 i_2 j_1 j_2} \epsilon^{i_1 \ldots i_n}\, \epsilon^{j_1 \ldots j_n} g_{i_3 j_3} \cdots g_{i_{n-1} j_{n-1}} \mu_{i_n}^n\, \mu_{j_n}^n / 4(n-3)!$$

or upon simplifying

$$K g^{ij}\, \mu_i^n\, \mu_j^n = {}^1/_4\, g^{i_1 i_2 i_3;\, j_1 j_2 j_3}\, R_{i_1 i_2 j_1 j_2}\, \mu_{i_3}^n\, \mu_{j_3}^n; \tag{14}$$

so that if K is to be independent of orientation we must have

$$2K\, g^{kl} = {}^1/_4\, (g^{i_1 i_2 k;\, j_1 j_2 l} + g^{i_1 i_2 l;\, j_1 j_2 k})\, R_{i_1 i_2 j_1 j_2}$$

which gives

$$2K\, g^{kl} = R\, g^{kl} - 2R^{kl}. \tag{15}$$

From this we get by contraction $R = \dfrac{2n}{n-2} K$ and

$$R_{kl} = \frac{2K}{n-1}\, g_{kl} = \frac{R}{n}\, g_{kl}. \tag{16}$$

But equations (16) define an Einstein space and by a theorem of Herglotz[7] we know that R is constant and consequently $K_{(n-1)}$ is. Thus the proof of Theorem III is completed.

Since if (16) are true it follows from (15) that $K_{(n-1)}$ is independent of orientation we may characterize an Einstein space geometrically by the theorem

IV. *A necessary and sufficient condition that a V_n be an Einstein space is that the curvature of all hyperplanes through any point be the same. Then this curvature does not vary from point to point.*

The word "all" in IV may be weakened to "n independent" as may be easily shown by means of (14).

We may now also state the inverse of Theorem II.

V. *If a space V_n has constant r-dimensional curvature for any one value of $r < n-1$, it has constant Riemannian curvature of every dimensionality.*

[1] Cf. M. S. Knebelman, *Proc. Nat. Acad. Sci.*, **16**, No. 2 (1930), pp. 156–159.

[2] Cf. L. P. Eisenhart, *Riemannian Geometry*, Princeton Univ. Press (1926).

[3] Cf. E. Cartan, *Géométrie des Espaces de Riemann*, Gauthier-Villars (1928).

[4] Cf. F. Schur, *Mathemat. Ann.*, **27** (1886), pp. 537–567.

[5] Cf. Wm. Flexner, *Ann. Math.*, **29**, No. 3 (1928), pp. 373–376.

[6] Cf. Oswald Veblen, "Invariants of Quadratic Differential Forms" (Cambridge Tract, No. 24).

[7] Cf. G. Herglotz, *Sitzungsber. Sächs. Gerellsch. Wiss.*, **68** (1916), pp. 769–822.

ON THE UNIFIED FIELD THEORY. III

By Tracy Yerkes Thomas

Department of Mathematics, Princeton University

Communicated December 9, 1930

It is the object of the present note[1] to deduce the general existence theorem of the Cauchy-Kowalewsky type for the system of field equations of the unified field theory. The existence theorem at which we arrive differs, however, from the usual form of such theorems in that certain linear combinations of partial derivatives of the h^i_α, namely, the absolute electromagnetic forces $h^i_{j,k}$, occur among the arbitrary functions, rather than these derivatives themselves. The theory is carried sufficiently far to determine the *characteristic surfaces* of the four-dimensional world, i.e., those three-dimensional surfaces which are characterized by the fact that if taken to "bear" the data of the problem our existence theorem ceases to apply, inasmuch as certain of the coefficients of the power series expansions of the h^i_α become indeterminant. These surfaces are analogous to the characteristic surfaces which arise in connection with the well-known equation

$$\Delta\varphi - \frac{1}{c^2}\frac{\partial^2\varphi}{\partial t^2} = 0,$$

which represents the propagation of a disturbance with constant velocity c; we have not, however, undertaken a discussion of the wave surfaces in the present note.

Before proceeding to the investigation of the mathematical problem at hand I would like to make a remark concerning the general nature of the field theory previously presented. Einstein[2] has pointed out that the vanishing of the invariant $h^i_{j,k}$ is the condition for the four-dimensional world to be Euclidean, or more properly pseudo-Euclidean. From the point of view of our previous notes this fact has its interpretation in the statement that the world will be pseudo-Euclidean only in the absence of electric and magnetic forces. This means that gravitational and electromagnetic phenomena must be intimately related since the existence of gravitation becomes dependent on the electromagnetic field. Thus we secure a real *physical unification* of gravitation and electricity in the sense that these concepts become but different manifestations of the same fundamental entity—provided, of course, that the theory shows itself to be tenable as a theory in agreement with experience.

Equations of preceding notes will be referred to by prefixing the serial number of the note to the number of the equation in question.

1. In the statement of the existence theorem which we shall deduce

the components h^i_α are divided into groups which exhibit different degrees of arbitrariness; for the purpose of exposition we shall presuppose this fact by dividing these components into suitable groups in accordance with the following

RULE: *The group* G_σ $(\sigma = 0, 1, 2, 3)$ *for the components* h^i_α *is composed of all components that can be formed from* h^i_α *by taking* $\alpha = \sigma + 1$ *and* $i = 1, 2, 3, 4$. There are four components h^i_α in G_σ and we shall denote these components by $P_{i\sigma}$ when reference is had to their division into groups. In an analogous manner the following rule will be selected to divide the components $h^i_{j,k}$ into groups.

RULE: *The group* $\mathfrak{G}_m (m = 0, 1, 2)$ *for the components* $h^i_{j,k}$ *is composed of all components that can be formed from* $h^i_{j,k}$ *by taking* $k = m + 1$ *and* $i, j = 1, 2, 3, 4$ *subject to the inequality* $j > m + 1$. If there are K_m components $h^i_{j,k}$ in group $\mathfrak{G}_m$, then

$$K_m = 12 - 4m.$$

We shall denote the components $h^i_{j,k}$ in group $\mathfrak{G}_m$ by K_{lm}, where $l = 1, \ldots, K_m$. It should be observed that the groups G and $\mathfrak{G}$ are mutually exclusive; also that the elements K_{lm} in any group $\mathfrak{G}_m$ are independent, i.e., unrelated by equations of the type I (4.10).

2. In terms of the notation introduced in Sect. 1 equations I (4.5) can be written

$$\frac{\partial P_{i\sigma}}{\partial x^\tau} = \frac{\partial P_{j\mu}}{\partial x^\nu} + \sum P^2 K, \qquad (2.1)$$

$$\begin{pmatrix} i = 1, 2, 3, 4 \\ \sigma = 1, 2, 3 \\ \tau = 1, \ldots, \sigma \end{pmatrix}$$

where the $\sum$ denotes a homogeneous polynomial, quadratic in the $P_{i\sigma}$ and linear in the K_{lm}; also the inequalities $\mu < \sigma$, $\mu < \nu$ are satisfied by the indices in these equations.

Now consider equations II (2.4), i.e.,

$$h^i_{j,k,l} + h^i_{k,l,j} + h^i_{l,j,k} = 2[h^i_{m,j}\, h^m_{k,l} + h^i_{m,k}\, h^m_{l,j} + h^i_{m,l}\, h^m_{j,k}]. \qquad (2.2)$$

Putting $k = 2$ and $l = 1$ in (2.2), these equations show that $h^i_{j,2,1}$ for $j > 2$ can be expressed in terms of derivatives of the K_{lo} plus a quadratic polynomial in the K_{lm}. If the contravariant vector component $h^1_1 \neq 0$ and if $R(P)$ denotes a rational function of the potentials $P_{i\sigma}$, these latter equations can be given the form

$$\frac{\partial K_{l1}}{\partial x^1} = \sum R(P) \frac{\partial K_{pq}}{\partial x^\nu} + \sum K^2,$$

where $l = 1. \ldots, 8$ and the inequalities $q \leqq 1$, $\nu > q$ are satisfied. More

generally, it can be observed that the system of equations (2.2) can be written

$$\frac{\partial K_{lm}}{\partial x^{\alpha}} = \sum R(P) \frac{\partial K_{pq}}{\partial x^{\nu}} + R(P) K^2, \qquad (2.3)$$

$$\begin{pmatrix} m = 1,2 \\ l = 1, \ldots, K_m \\ \alpha = 1, \ldots, m \end{pmatrix}$$

provided that the quantities

$$h_1^1 \text{ and } \begin{vmatrix} h_1^1 & h_2^1 \\ h_1^2 & h_2^2 \end{vmatrix}$$

constructed from the h_i^{σ} components do not vanish; the inequalities $q \leqq m$, $\nu > q$ are also satisfied by the indices in (2.3).

3. When account is taken of the field equations

$$\Delta_k h^i_{j,k} = 0 \qquad (3.1)$$

the form of (2.3) will remain unchanged, provided that an arrangement of the components $h^i_{j,k}$ into new groups $\mathfrak{G}^*_m$ is effected. To see how this is

$j\,k\,l$	$\frac{\partial h^i_{3,1}}{\partial x^1}$	$\frac{\partial h^i_{4,1}}{\partial x^1}$	$\frac{\partial h^i_{3,2}}{\partial x^1}$	$\frac{\partial h^i_{4,2}}{\partial x^1}$	$\frac{\partial h^i_{2,1}}{\partial x^1}$	$\frac{\partial h^i_{2,1}}{\partial x^2}$	$\frac{\partial h^i_{4,3}}{\partial x^1}$	$\frac{\partial h^i_{4,3}}{\partial x^2}$
123	h^1_2	0	$-h^1_1$	0	$-h^1_3$	$-h^2_3$	0	0
124	0	h^1_2	0	$-h^1_1$	$-h^1_4$	$-h^2_4$	0	0
134	$-h^1_4$	h^1_3	0	0	0	0	$-h^1_1$	$-h^2_1$
234	0	0	$-h^1_4$	h^1_3	0	0	$-h^1_2$	$-h^2_2$
$j = 1$	h^1_3	h^1_4	0	0	h^1_2	h^2_2	0	0
$j = 2$	0	0	h^1_3	h^1_4	h^1_1	h^2_1	0	0
$j = 3$	h^1_1	0	$-h^1_2$	0	0	0	h^1_4	h^2_4
$j = 4$	0	h^1_1	0	$-h^1_2$	0	0	$-h^1_3$	$-h^2_3$

TABLE 1

brought about we consider the four sets of equations (2.2) corresponding to $j\,k\,l$ equal 123, 124, 134 and 234, respectively; likewise the four sets of equations obtained by letting j have the values 1, 2, 3 and 4 in (3.1). The equations so obtained can be solved for the following derivatives

$$\frac{\partial h^i_{3,1}}{\partial x^1}, \frac{\partial h^i_{4,1}}{\partial x^1}, \frac{\partial h^i_{3,2}}{\partial x^1}, \frac{\partial h^i_{4,2}}{\partial x^1}, \frac{\partial h^i_{2,1}}{\partial x^1}, \frac{\partial h^i_{2,1}}{\partial x^2}, \frac{\partial h^i_{4,3}}{\partial x^1}, \frac{\partial h^i_{4,3}}{\partial x^2}.$$

We have, in fact, indicated the determinant formed by the coefficients of the above derivatives in Table 1 which is so constructed that (1) each row corresponds to an equation of the system, namely, an equation (2.2) or (3.1) determined by the indicated values of $j\ k\ l$ or j, respectively, and (2) the elements h_i^α in any column are the coefficients of the derivatives at the top of this column.

By an ingenious method J. M. H. Wedderburn[3] has shown that the determinant in Table 1 is equal to the product of two fourth-order conjugate imaginary determinants. This leads to the easy reduction of this determinant to the form

$$\left(\sum_{k=1}^{4} e_k\, h_k^1\, h_k^1\right)^2 \left\{\left|\begin{matrix} h_1^1 & h_1^2 \\ h_2^1 & h_2^2 \end{matrix}\right|^2 + \left|\begin{matrix} h_3^1 h_3^2 \\ h_4^1 h_4^2 \end{matrix}\right|^2\right\}. \tag{3.2}$$

I shall take this opportunity to express my thanks to Professor Wedderburn for this contribution.

Let us now denote the components $h_{j,k}^i$ in the four sets of derivatives in the first four columns of Table 1 by K_{l1}^* and the components $h_{j,k}^i$ in the two sets of derivatives in the last four columns of the table by K_{l2}^*; we indicate this readily by the following schematic arrangement

$$K_{l1}^* \sim h_{3,1}^i,\ h_{4,1}^i,\ h_{3,2}^i,\ h_{4,2}^i$$
$$K_{l2}^* \sim h_{2,1}^i,\ h_{4,3}^i.$$

All independent $h_{j,k}^i$ are included in the components K^*. There are $K_1^* = 16$ components K_{l1}^*, and $K_2^* = 8$ components K_{l2}^*, to introduce a notation corresponding to that employed in Sect. 2. In terms of these designations it is therefore clear that if the expression (3.2) is different from zero, equations (2.2) and (3.1) can be put into the form

$$\frac{\partial K_{lm}^*}{\partial x^\alpha} = \sum R\,(P)\,\frac{\partial K_{pq}^*}{\partial x^\nu} + \sum R\,(P)\,K^{*2}, \tag{3.3}$$

$$\begin{pmatrix} m &= 1, 2 \\ l &= 1, \ldots, K^* \\ \alpha &= 1, \ldots, m \end{pmatrix}$$

where the inequality $\nu > q$ is satisfied by the indices in these equations.

4. The system composed of (2.1) and (3.3) is harmonic in the sense of Riquier.[4] Moreover, this system is completely integrable. To see this let us observe that if we form the conditions of integrability we obtain a system R involving the following quantities

$$P_{j\mu},\ \frac{\partial P_{j\mu}}{\partial x^\nu},\ \frac{\partial^2 P_{j\mu}}{\partial x^\nu\, \partial x^\xi} \tag{4.1a}$$

$$\begin{pmatrix} j = 1, 2, 3, 4; & \mu = 0, 1, 2, 3 \\ \nu, \xi = 1, 2, 3, 4; & \nu \geqq \xi; \quad \nu, \xi > \mu \end{pmatrix}$$

and

$$K^*_{pq}, \quad \frac{\partial K^*_{pq}}{\partial x^\nu}, \quad \frac{\partial^2 K^*_{pq}}{\partial x^\nu \partial x^\xi}. \tag{4.1b}$$

$$\begin{pmatrix} q = 1, 2; & p = 1, \ldots, K^*_q \\ \nu, \xi = 2, 3, 4; & \nu \geqq \xi; \quad \nu, \xi > q \end{pmatrix}$$

Now it can be shown that the components h^i_α and also that, subject to the condition of symmetry in their lower indices, the components $\Lambda^\alpha_{\beta\gamma}$ and $\Lambda^\alpha_{\beta\gamma\delta}$ can be assigned arbitrary values at a point Q of the space;[5] moreover, arbitrary values can be assigned to these quantities at Q irrespective of the values at Q of the quantities in (4.1b). The number of arbitrary quantities h^i_α, $\Lambda^\alpha_{\beta\gamma}$ and $\Lambda^\alpha_{\beta\gamma\delta}$ is therefore equal to 136, and since these quantities are determined by the quantities in (4.1a), it follows that there can not be less than 136 of the quantities in (4.1a) to which arbitrary values can be assigned at the point Q. There are also 24 arbitrary quantities K^*_{pq} in (4.1b). Now in Note II we established the expression

$$16K(3, r + 1) + 8K(2, r + 1)$$

as a lower bound to the number of arbitrary derivatives of the $(r + 1)st$ order of the quantities $h^i_{j,k}$. We deduce from this that no less than 184 of the derivatives of the K^*_{pq} in (4.1b) can have arbitrary values at Q. Hence 136 + 24 + 184, or 344, is a lower bound to the number of arbitrary quantities in (4.1). Now 344 is also the number of arbitrary quantities in (4.1) as deduced by actual calculation based on the admissible ranges of the indices of these quantities. It follows, therefore, that all quantities in (4.1) can be given arbitrary values at a point Q of space and hence that the above system R, giving the conditions of integrability of (2.1) and (3.3), must be satisfied identically. This gives the following

EXISTENCE THEOREM: *Let*

$$\begin{array}{c|c} \mathfrak{P}_{i\sigma}(x^{\sigma+1}, x^4) & \mathfrak{R}^*_{lm}(x^{m+1}, \ldots, x^4) \\ [i = 1, 2, 3, 4; \ \sigma = 0, 1, 2, 3] & [m = 1, 2; \ l = 1, \ldots, K^*_m] \end{array}$$

denote sets of functions of the variables $x^{\sigma+1}, \ldots, x^4$ and $x^{m+1}, \ldots, x^4$, respectively, analytic in the neighborhood of the values $x^\alpha = q^\alpha$ of their arguments; furthermore the $\mathfrak{P}_{i\sigma}$ are subject to the condition that at $x^\alpha = q^\alpha$ the corresponding determinant $|h^i_\alpha|$ and the expression (3.2) *do not vanish. Then there exists one, and only one, set of potentials h^i_α in a system of coordinates x^α, each function $h^i_\alpha(x)$ being analytic in the neighborhood of the values $x^\alpha = q^\alpha$, which constitutes a set of integrals of the field equations* (3.1) *such that*

$$P_{io} = \mathfrak{P}_{io}(x^1,x^2,x^3,x^4) \quad [i = 1, 2, 3, 4] \;\Big|\; P_{i\sigma} = \mathfrak{P}_{i\sigma}(x^{\sigma+1}, \ldots, x^4) \quad \begin{bmatrix} i = 1,2,3,4; \sigma = 1,2,3 \\ x^1 = q^1, \ldots, x^\sigma = q^\sigma \end{bmatrix} \;\Big|\; K^*_{lm} = \mathfrak{K}^*_{lm}(x^{m+1}, \ldots, x^4). \quad \begin{bmatrix} m = 1,2; l = 1, \ldots, K^*_m \\ x^1 = q^1, \ldots, x^m = q^m \end{bmatrix}$$

5. Let us denote by S_k a k dimensional surface in the four dimensional space S_4, and consider the three surfaces S_1, S_2, S_3 defined by

$$\begin{cases} S_3: \; x^1 = \varphi(x^2, x^3, x^4) \\ S_2: \begin{cases} x^1 = \varphi(x^2, x^3, x^4) \\ x^2 = \psi(x^3, x^4) \end{cases} \\ S_1: \begin{cases} x^1 = \varphi(x^2, x^3, x^4) \\ x^2 = \psi(x^3, x^4) \\ \omega^3 = \omega(x^4), \end{cases} \end{cases} \tag{5.1}$$

where the functions φ, ψ, ω are analytic in the neighborhood of some set of values $x^i = q^i$ of the coördinates; as so defined any surface S_i is contained within the surface S_k provided that the inequality $i<k$ is satisfied. We shall take the above surfaces S_1, S_2, S_3 and the surface S_4, or four dimensional space itself, as those surfaces which "bear" the data in our problem. In other words, we now assign the $P_{i\sigma}$ over the surface $S_{4-\sigma}$ and the K^*_{lm} over the surface S_{4-m} as shown by the scheme

$$\begin{array}{c|c} \text{Over } S_{4-\sigma}: & \text{Over } S_{4-m}: \\ P_{i\sigma} = \mathfrak{P}_{i\sigma}(x^{\sigma+1}, \ldots, x^4) & K^*_{lm} = \mathfrak{K}^*_{lm}(x^{m+1}, \ldots, x^4), \\ [i = 1, 2, 3, 4; \; \sigma = 0, 1, 2, 3] & [m = 1, 2; \; l = 1, \ldots, K^*_m] \end{array} \tag{5.2}$$

where the $\mathfrak{P}$ and $\mathfrak{K}^*$ are analytic functions of the indicated surface coordinates.

Now let q^i denote the coördinates of a point on the surface S_1 and make the transformation

$$y^1 = x^1 - \varphi, \; y^2 = x^2 - \psi, \; y^3 = x^3 - \omega, \; y^4 = x^4 - q^4 \tag{5.3}$$

so that the above surfaces S_1, S_2, S_3 become

$$\begin{cases} S_3: \; y^1 = 0 \\ S_2: \begin{cases} y^1 = 0 \\ y^2 = 0 \end{cases} \\ S_1: \begin{cases} y^1 = 0 \\ y^2 = 0 \\ y^3 = 0. \end{cases} \end{cases} \tag{5.4}$$

In order that the existence theorem of Sect. 4 may apply it is necessary that the data given for the x coördinate system be determined with respect to the y coördinate system over the surfaces S. It is obvious that the

absolute electromagnetic forces will be so determined since these quantities, i. e., K^*_{lm}, are absolute invariants under coördinate transformations. If we denote the components h^i_α by l^i_α with respect to the y coördinate system, then the components l^i_α in group G_σ are likewise determined over the surface $S_{4-\sigma}$. In connection with the demonstration of this fact we lay down the following:

DEFINITION: *A group of components G which is related by a transformation T to the corresponding components in groups G of the same and lower serial order, will be said to be closed with respect to the transformation T.* Since the groups G_σ are closed with respect to the transformation (5.3) in this sense, the above statement, namely, that the components l^i_α in group G_σ are determined over the surface $S_{4-\sigma}$, is immediately evident.

Let us now put

$$\Phi = x^1 - \varphi(x^2, x^3, x^4), \quad \Psi = x^2 - \psi(x^3, x^4)$$
$$\Omega = x^3 - \omega(x^4), \quad \Xi = x^4 - q^4$$

and consider the equation

$$\boxed{g^{\alpha\beta} \frac{\partial \Phi}{\partial x^\alpha} \frac{\partial \Phi}{\partial x^\beta} = 0} \tag{5.5}$$

in which the $g^{\alpha\beta}$ denote the ordinary contravariant components of the fundamental metric tensor, i.e.,

$$g^{\alpha\beta} = \sum_{k=1}^{4} e_k h^\alpha_k h^\beta_k.$$

Under the transformation (5.3) the left member of (5.5) becomes an expression in the l^i_α having the form of the first factor of (3.2); similarly the left members of the two equations

$$\boxed{\sum_{\alpha=1}^{4} \sum_{\beta=1}^{4} \begin{vmatrix} h^\alpha_1 & h^\beta_1 \\ h^\alpha_2 & h^\beta_2 \end{vmatrix} \frac{\partial \Phi}{\partial x^\alpha} \frac{\partial \Psi}{\partial x^\beta} = 0} \tag{5.6a}$$

and

$$\boxed{\sum_{\alpha=1}^{4} \sum_{\beta=1}^{4} \begin{vmatrix} h^\alpha_3 & h^\beta_3 \\ h^\alpha_4 & h^\beta_4 \end{vmatrix} \frac{\partial \Phi}{\partial x^\alpha} \frac{\partial \Psi}{\partial x^\beta} = 0} \tag{5.6b}$$

assume a form in the components l^i_α corresponding to the first and second determinants in the last factor of the expression (3.2). Suppose that the equation (5.5) and at least one of the equations (5.6), i.e. (5.6a) or (5.6b), are not satisfied over the surface S_1; then take the above-mentioned point on S_1 with coördinate q^4 occurring in the transformation (5.3) to be a point on S_1 at which the left member of (5.5) and the left member of either (5.6a) or (5.6b) do not vanish. The existence theorem of Sect. 4 will then apply with respect to the y coördinate system and on the basis of this application the following existence theorem can be stated.

EXISTENCE THEOREM: *Consider a set of three surfaces* S_1, S_2, S_3 *defined by* (5.1) *and a set of analytic functions* $\mathfrak{P}$ *and* $\mathfrak{R}^*$ *defined over these surfaces and the four dimensional space* S_4 *as represented by* (5.2), *the functions* $\mathfrak{P}$ *being subject to the condition that* (1) *the corresponding determinant* $|h^i_\alpha|$ *does not vanish over* S_1, *and* (2) *the equation* (5.5) *and either equation* (5.6a) *or* (5.6b) *are not satisfied over the surface* S_1. *Then there exists one, and only one, set of potentials* h^i_α *in a system of coördinates* x^α, *the* h^i_α (x) *being analytic functions of the* x^α *coördinates, which constitutes a set of integrals of the field equations* (3.1) *such that the quantities* $P_{i\sigma}$ *and* K^*_{lm} *assume the assigned values* $\mathfrak{P}$ *and* $\mathfrak{R}^*$ *over the surfaces* S.

6. Equations (5.5) and (5.6a) satisfied over a curve S_1 are invariant under coördinate transformations. It is possible, however, to make an orthogonal transformation of the fundamental vectors, i.e., a transformation of the form

$$*h^i_\alpha = a^i_k h^k_\alpha , \tag{6.1}$$

so that with respect to the $*h$ potentials an equation of the type (5.6a) will not be satisfied over the curve S_1.[6] Equations (5.5), on the other hand, remain invariant in form under an orthogonal transformation (6.1) of the fundamental vectors. A curve S_1 over which the equation (5.5) is satisfied, will be called a *characteristic line with respect to the surface* S_3. Those surfaces S_3 over which the equation (5.5) is satisfied will be called *characteristic surfaces;* they are analogous to the characteristic surfaces determined by the wave equation as mentioned in the introduction to this note, and must therefore be expected to represent the wave surfaces in the present theory.

By a formal process based on the general theory of characteristic surfaces as developed by Volterra and Hadamard, it was recently shown by Levi-Civita[7] that an equation of the form (5.5) determines the characteristic surfaces for the Einstein gravitational equations in regions free of matter; the existence theorem for the equations of Einstein, however, was not established by Levi-Civita. An investigation of the characteristic surfaces determined by the field equations (3.1) will be undertaken in a later communication.

[1] Previous notes on the Unified Field Theory have appeared in these PROCEEDINGS, **16**, 761–776 and 830–835 (1930).

[2] A. Einstein, "Die Kompatibilität der Feldgleichungen in der einheitlichen Feldtheorie," *Sitzungsberichte preussischen Akad. Wissenschaften*, 1930, 18–23.

[3] Wedderburn's method is as follows: For simplicity the revised notation

$$\alpha = h^1_1,\ \beta = h^1_2,\ \gamma = h^1_3,\ \delta = h^1_4$$
$$a = h^2_1,\ b = h^2_2,\ c = h^2_3,\ d = h^2_4$$

is introduced. Then (1) if we rearrange the rows and columns, and (2) if we multiply by -1 certain of the rows and columns in Table 1, this table assumes the form represented

by Table 2. Denoting $\sqrt{-1}$ by i as usual, we now add i times each even column to the preceding odd column in Table 2 and then subtract i times each odd row from the preceding even row. Table 3 shows the result obtained. By an obvious rearrangement of rows and columns Table 3 assumes the form given in Table 4, which shows that the determinant of the elements in Table 1 is equal, apart from algebraic sign which we have disregarded, to the product of two-fourth order conjugate imaginary determinants. Expansion of the fourth-order determinant in the lower right hand corner of Table 4 shows that this determinant has the value

$$(\alpha^2 - \beta^2 - \gamma^2 - \delta^2)[(a\beta - b\alpha) + i(c\delta - d\gamma)].$$

Taking the absolute value of this expression, the expression (3.2) is obtained. The fact that (3.2) and the determinant formed from the elements in Table 1 have the same algebraic sign is easily observed by replacing the h_i^α by Kronecker's δ_i^α in (3.2) and Table 1.

0	$-c$	0	$-\gamma$	α	0	$-\beta$	0
c	0	γ	0	0	α	0	$-\beta$
d	0	δ	0	$-\beta$	0	α	0
0	d	0	δ	0	$-\beta$	0	α
a	0	α	0	0	γ	δ	0
0	a	0	α	$-\gamma$	0	0	δ
b	0	β	0	δ	0	0	γ
0	b	0	β	0	δ	$-\gamma$	0

TABLE 2

$-ie$	$-e$	$-i\gamma$	$-\gamma$	α	0	$-\beta$	0
0	ic	0	$i\gamma$	0	α	0	$-\beta$
d	0	δ	0	$-\beta$	0	α	0
0	d	0	δ	0	$-\beta$	0	α
a	0	α	0	$i\gamma$	γ	δ	0
0	a	0	α	0	$-i\gamma$	0	δ
b	0	β	0	δ	0	$i\gamma$	γ
0	b	0	β	0	δ	0	$-i\gamma$

TABLE 3

$-ic$	$-i\gamma$	α	$-\beta$	$-c$	$-\gamma$	0	0
d	δ	$-\beta$	α	0	0	0	0
a	α	$i\gamma$	δ	0	0	γ	0
b	β	δ	$i\gamma$	0	0	0	γ
0	0	0	0	ic	$i\gamma$	α	$-\beta$
0	0	0	0	d	δ	$-\beta$	α
0	0	0	0	a	α	$-i\gamma$	δ
0	0	0	0	b	β	δ	$-i\gamma$

TABLE 4

[4] C. Riquier, "De l'existence des intégrales dans un système differential quelconque," *Annales de l'ecole normale superieure, serie* 3, **10** (1893), pp. 65–86; *Ibid.*, 123–50. Two of the three conditions in Riquier's definition of the harmonic system are obviously

satisfied for the system composed of (2.1) and (3.3). To show that the third condition is likewise satisfied we assign to each of the quantities x^α, $P_{i\sigma}$, and K^*_{lm} a single "cote" in the following manner:

x^α has the "cote" $-\alpha$,

$P_{i\sigma}$ has the "cote" σ,

K^*_{lm} has the "cote" m.

Then $(\sigma - \tau)$ and $(m - \alpha)$ are the "cotes" of the derivatives in the left members of the equations (2.1) and (3.3), respectively; similarly $(m - \alpha)$ and $(q - \nu)$ are the "cotes" of derivatives in the right members of these equations. Hence

$$(\sigma - \tau) - (\mu - \nu) > 0, \quad (m - \alpha) - (q - \nu) > 0,$$

since $\sigma - \tau \geqq 0$, $m - \alpha \geqq 0$, $\mu - \nu < 0$, and $q - \nu < 0$. The system composed of (2.1) and (3.3) is therefore harmonic in the sense of Riquier. See also T. Y. Thomas, "Invariantive Systems of Partial Differential Equations," *Ann. Math.*, **31** (1930), 687–713; and *Ibid.*, 714–726.

[5] For an analogous proof, see T. Y. Thomas, "A Theorem Concerning the Affine Connection," *Am. J. Math.*, **50** (1928), 518–520.

[6] We have

$$({}^*h^\alpha_1\, {}^*h^\beta_2 - {}^*h^\alpha_2\, {}^*h^\beta_1)\frac{\partial\Phi}{\partial x^\alpha}\frac{\partial\Psi}{\partial x^\beta} = \left[(h^\alpha_k\, h^\beta_l - h^\alpha_l\, h^\beta_k)\frac{\partial\Phi}{\partial x^\alpha}\frac{\partial\Psi}{\partial x^\beta}\right] a^k_1\, a^l_2, \qquad (a)$$

in which the expression in brackets in the right member is assumed to be taken at a point P on S_1. Now assume that the left member of (a) is linearly dependent on the independent polynomials of the set

$$\sum_{k=1}^{4} e_k\, a^k_p\, a^k_q - e_p\, \delta^p_q. \qquad (b)$$

Then we can write

$$\left[(h^\alpha_k\, h^\beta_l - h^\alpha_l\, h^\beta_k)\frac{\partial\Phi}{\partial x^\alpha}\frac{\partial\Psi}{\partial x^\beta}\right] a^k_1\, a^l_2 = \sum_{k=1}^{4} (A^{pq}\, e_k\, \delta^k_l)\, a^k_p\, a^l_q - \sum_{k=1}^{4} e_k\, A^{kk},$$

where the A^{pq} are constants. Equating coefficients, we obtain

$$(h^\alpha_k\, h^\beta_l - h^\alpha_l\, h^\beta_k)\frac{\partial\Phi}{\partial x^\alpha}\frac{\partial\Psi}{\partial x^\beta} = A e_k\, \delta^k_l,$$

where A is a constant. Putting $k = l$ in this last equation we see that A must vanish. Hence we can deduce that

$$\frac{\partial\Phi}{\partial x^\alpha}\frac{\partial\Psi}{\partial x^\beta} - \frac{\partial\Phi}{\partial x^\beta}\frac{\partial\Psi}{\partial x^\alpha} = 0.$$

In other words the rank of the matrix

$$\left\| \begin{matrix} \frac{\partial\Phi}{\partial x^1} & \frac{\partial\Phi}{\partial x^2} & \frac{\partial\Phi}{\partial x^3} & \frac{\partial\Phi}{\partial x^4} \\ \frac{\partial\Psi}{\partial x^1} & \frac{\partial\Psi}{\partial x^2} & \frac{\partial\Psi}{\partial x^3} & \frac{\partial\Psi}{\partial x^4} \end{matrix} \right\|$$

is less than two at P. However, this is not possible since the above matrix is seen to have the form

$$\left\| \begin{matrix} 1 & * & * & * \\ 0 & 1 & * & * \end{matrix} \right\|$$

when account is taken of the functions Φ and ψ in Sect. 5. Hence at a point P on S_1 the right member of (a) is linearly independent of the independent polynomials of the set (b). It is therefore possible to choose the values of the constants a_k^i so that at a point P on S_1 the right member of (a) and the independent polynomials (b) can have independent values; in particular, the constants a_k^i can be so chosen that the right member of (a) will be different from zero while all expressions (b) vanish. This proves the statement in Sect. 6 that by making a suitable orthogonal transformation (6.1) of the fundamental vectors, the expression in the left member of (a) can be given a value different from zero over the curve S_1.

[7] T. Levi-Civita, "Caratteristiche e bicaratteristiche delle equazioni gravitazionali di Einstein," *Rendiconti Accad. Lincei*, **11**, 1930, pp. 1–11; *Ibid.*, pp. 113–121. In this connection, see also T. Y. Thomas, "On the Existence of Integrals of Einstein's Gravitational Equations for Free Space and Their Extension to n Variables," these PROCEEDINGS, **15** (1929), pp. 906–913.

QUANTUM-MECHANICAL MOTION OF FREE ELECTRONS IN ELECTROMAGNETIC FIELDS

BY E. H. KENNARD

DEPARTMENT OF PHYSICS, CORNELL UNIVERSITY

Communicated November 25, 1930

The motion of an electron in an electric or magnetic field was treated quantum-mechanically by the author and by Darwin for the case of a certain type of wave-packet in a uniform field.[1,2] Later writers[3,4,5] have concerned themselves for the most part with the energy and the characteristic functions, which are needed for radiation problems. In this paper formulas are obtained by the Ehrenfest method for the motion of the centroid of any packet in an electromagnetic field of general type, the field being treated in the usual way as a perturbing term in the Hamiltonian function for the electron. Only the non-relativistic case is considered, based upon Schrödinger's equation; for the Dirac electron the usual method of approximation leads to this same equation plus terms expressing the spin effect,[4] but the attempt to justify the approximation for such purposes encounters an extraordinary difficulty that seems to be connected with the "negative energy" problem and has not yet been overcome.

The general result obtained below is that in a uniform electric and magnetic field the packet-centroid moves as the electron should move according to classical theory, while in non-uniform fields its acceleration is a kind of average of the classical value.

The Hamiltonian for an electron in an electromagnetic field, in which for generality we shall include an extraneous potential-energy function $V(x, t)$, can be written:

$$H = \frac{1}{2m}\left(\mathbf{p} - \frac{e}{c}\mathbf{A}\right)\cdot\left(\mathbf{p} - \frac{e}{c}\mathbf{A}\right) + e\varphi + V, \tag{1}$$

where e = algebraic charge, m = mass and $\mathbf{p}$ = vector momentum of the electron, $\mathbf{A}$ = vector and φ = scalar potential of the field, and c = speed of light. In this form H is "symmetrized," i.e., it retains its value when turned end for end with reversal of the order of all factors; such a form is necessary in order to secure harmony with classical theory on the macroscopic scale, except when the field satisfies everywhere the condition, div $\mathbf{A} = -(1/c)(\partial\varphi/\partial t) = 0$. We then have as wave-equation:

$$-\epsilon\frac{\partial\psi}{\partial t} = H\left(\epsilon\frac{\partial}{\partial q}, q, t\right)\psi = \frac{\epsilon^2}{2m}\Delta\psi - \frac{e}{2mc}[\text{div}\,(\mathbf{A}\psi) + \mathbf{A}\cdot\nabla\psi] + e\varphi\psi + V\psi, \tag{2}$$

n which $\epsilon = h/2\pi i$.

The coördinates of the electron being denoted by x_k (or simply x in $V(x, t)$, etc.), those of the "centroid" of the wave-packet are defined by

$$\bar{x}_k = \int \psi^* x_k \psi dx, \qquad k = 1, 2, 3, \tag{3}$$

where $dx = dx_1dx_2dx_3$ and ψ^* is the complex conjugate of ψ. The value of $d^2\bar{x}_k/dt^2$ can be calculated by a two-fold application of (2), but the work gains greatly in brevity and clarity if we first collect a few logarithms for integrals such as (3) and then proceed algebraically until the end is reached.

In general, if $f(p, q, t)$ is any measurable function of the coördinates, momenta and time, its mean expectation of value is

$$\overline{f(p, q, t)} = \int \psi^* f\left(\epsilon\frac{\partial}{\partial q}, q, t\right)\psi dq. \tag{4}$$

Clearly if $g(p, q, t)$ is any other such function, or if b is a constant, then

$$\overline{f + g} = \bar{f} + \bar{g}, \quad \overline{bf} = b\bar{f}. \tag{5}$$

For the time derivative we have, using Schrödinger's equation in terms of the Hamiltonian $H(p, q, t)$,

$$\frac{d}{dt}\overline{f(p, q, t)} = \int\left(\psi^* f\frac{\partial\psi}{\partial t} + \frac{\partial\psi^*}{\partial t}f\psi + \psi^*\frac{\partial f}{\partial t}\psi\right)dq = \int\psi^*\frac{\partial f}{\partial t}\psi dq.$$

$$-\frac{1}{\epsilon}\int\psi^* f H\left(\epsilon\frac{\partial}{\partial q}, q, t\right)\psi dq + \frac{1}{\epsilon}\int\left[H\left(-\epsilon\frac{\partial}{\partial q}, q, t\right)\psi^*\right]f\psi dq.$$

In the last integral let us "roll" ("wälzen") H off of ψ^* by integrating by parts and rejecting each time the integrated part. Being symmetrized, H is thereby unaltered in form, but the repeated changes of sign result in replacing $-\epsilon\partial/\partial q$ by $+\epsilon\partial/\partial q$, so that we find:

$$\frac{d}{dt}\overline{f(p, q, t)} = \frac{1}{\epsilon}\int \psi^*\left[H\left(\epsilon\frac{\partial}{\partial q}, q, t\right)f\left(\epsilon\frac{\partial}{\partial q}, q, t\right) - f\left(\epsilon\frac{\partial}{\partial q}, q, t\right) H\left(\epsilon\frac{\partial}{\partial q}, q, t\right)\right]\psi dq + \int \psi^*\frac{\partial f}{\partial t}\left(\epsilon\frac{\partial}{\partial q}, q, t\right)\psi dq. \quad (6)$$

The bracket in the first integral in (6) we now recognize as a familiar bit of q-number or matrix theory; we can evaluate it in the usual way, before forming the differential operators, by an algebraic manipulation in which we treat the p's and q's as variables that fail to commute only when a p meets its own q, in which case

$$p_iq_i - q_ip_i = \epsilon, \quad \text{since} \quad \frac{\partial}{\partial q_i}q_i - q_i\frac{\partial}{\partial q_i} = 1. \quad (7)$$

In this way one finds, upon substituting q_i, p_i and $\eta(q, t)$ in turn for f:

$$\frac{d}{dt}\overline{q_i} = \overline{\frac{\partial H}{\partial p_i}}, \quad \frac{d}{dt}\overline{p_i} = -\overline{\frac{\partial H}{\partial q_i}}, \quad (8)$$

$$\frac{d}{dt}\overline{\eta(q, t)} = \sum(\sigma)\overline{\left(\frac{\partial \eta}{\partial q_\sigma}\frac{\partial}{\partial p_\sigma}\right)H} + \overline{\frac{\partial \eta}{\partial t}}, \quad (9)$$

in which $\left(\frac{\partial \eta}{\partial q_\sigma}\frac{\partial}{\partial p_\sigma}\right)H$ is to be calculated for each term of H by replacing each p_σ^1 in it in succession by $\partial\eta/\partial q_\sigma$ and adding the results, while the first derivative in (8) is calculated by replacing each p_i^1 by 1.

Equations (8) throw into vivid relief the approximate agreement of wave and classical mechanics for cases in which we can neglect the finite size of the packet and deal only with the averages.

Substituting now x_k for q_i and H from (1) in (8) we have:

$$\frac{d}{dt}\overline{x_k} = \frac{1}{m}\overline{p_k} - \frac{e}{cm}\overline{A_k},$$

$$\frac{d}{dt}\overline{p_k} = -e\frac{\partial \varphi}{\partial x_k} - \frac{\partial V}{\partial x_k} + \frac{e}{2cm}\left[\frac{\partial \mathbf{A}}{\partial x_k}\cdot\left(\mathbf{p} - \frac{e}{c}\mathbf{A}\right) + \left(\mathbf{p} - \frac{e}{c}\mathbf{A}\right)\cdot\frac{\partial \mathbf{A}}{\partial x_k}\right],$$

where in the second equation, as likewise in the following three, we have for simplicity left the averaging-bar to be understood as drawn over the entire right-hand member. Replacing η in (9) by $A_k(x, t)$ we find also

$$\frac{d}{dt}\overline{A_k} = \frac{1}{2m}\left[(\nabla A_k)\cdot\left(\mathbf{p} - \frac{e}{c}\mathbf{A}\right) + \left(\mathbf{p} - \frac{e}{c}\mathbf{A}\right)\cdot(\nabla A_k)\right] + \frac{\partial A_k}{\partial t}.$$

Then $\frac{d^2}{dt^2}\overline{x_k} = \frac{1}{m}\frac{d}{dt}\overline{p_k} - \frac{e}{cm}\frac{d}{dt}\overline{A_k} = -\frac{e}{m}\frac{\partial \varphi}{\partial x_k} - \frac{1}{m}\frac{\partial V}{\partial x_k} - \frac{e}{cm}\frac{\partial A_k}{\partial t} +$

$$\frac{e}{2cm^2}\left[\left(\frac{\partial \mathbf{A}}{\partial x_k} - \nabla A_k\right)\cdot\left(\mathbf{p} - \frac{e}{c}\mathbf{A}\right) + \left(\mathbf{p} - \frac{e}{c}\mathbf{A}\right)\cdot\left(\frac{\partial \mathbf{A}}{\partial x_k} - \nabla A_k\right)\right].$$

Now the electric field-strength is $\mathbf{E} = -\nabla\varphi - (1/c)(\partial\mathbf{A}/\partial t)$ while the components of $\partial\mathbf{A}/\partial x_k - \nabla A_k$ reduce either to 0 or to components of the magnetic field-strength $\mathbf{H} = \text{curl } \mathbf{A}$, with or without a minus sign; and, in accordance with (7),

$$\mathbf{p}\cdot\left(\frac{\partial \mathbf{A}}{\partial x_k} - \nabla A_k\right) - \left(\frac{\partial \mathbf{A}}{\partial x_k} - \nabla A_k\right)\cdot\mathbf{p} = \epsilon \,\text{div}\left(\frac{\partial \mathbf{A}}{\partial x_k} - \nabla A_k\right).$$

We thus find, upon returning to the explicit integral form, in vector notation,

$$\frac{d}{dt}\bar{\mathbf{x}} = \bar{\mathbf{v}} = \int \psi^*\left(\frac{h}{2\pi i m}\nabla\psi - \frac{e}{cm}\mathbf{A}\psi\right)dx, \tag{10}$$

$$\frac{d^2}{dt^2}\bar{\mathbf{x}} = -\frac{1}{m}\int \psi^*\nabla' \psi dx + \frac{e}{m}\int \psi^*\mathbf{E}\psi dx$$

$$+ \frac{e}{cm}\int \psi^*\left[\left(\frac{h}{2\pi i m}\nabla\psi - \frac{e}{cm}\mathbf{A}\psi\right)\times\mathbf{H}\right]\psi dx$$

$$+ \frac{eh}{4\pi i c m^2}\int \psi^*(\nabla \text{ div } \mathbf{A} - \nabla^2\mathbf{A})\psi dx. \tag{11}$$

The last integral in the second of these equations is obviously imaginary; it merely serves to remove the imaginary part of the preceding one, as can be either verified directly or inferred from the obvious reality of all other terms in the equation. We might omit it and write simply:

$$\frac{d^2}{dt^2}\bar{\mathbf{x}} = -\frac{1}{m}\int \psi^*\nabla V\psi dx + \frac{e}{m}(R)\int \psi^*\left[\mathbf{E} + \frac{1}{c}\left(\frac{h}{2\pi i m}\frac{1}{\psi}\nabla\psi - \frac{e}{cm}\mathbf{A}\right)\times\mathbf{H}\right]\psi dx \tag{12}$$

where (R) means that only the real part of the integral following it is to be taken. The expression

$$\frac{h}{2\pi i m}\frac{1}{\psi}\nabla\psi - \frac{e}{cm}\mathbf{A}$$

can be interpreted as the "velocity of the probability," whose mean value is $\bar{\mathbf{v}}$; the second integral in (12) is then simply the mean value of the classical force per unit charge, $\mathbf{E} + \mathbf{v}\times\mathbf{H}/c$.

When the electric and magnetic fields are uniform we have further $\int\psi^*\mathbf{E}\psi dx = \mathbf{E}$, etc., so that

$$\frac{d^2}{dt^2}\bar{\mathbf{x}} = -\frac{1}{m}\int\psi^*\nabla V\psi dx + \frac{e}{m}\left(\mathbf{E} + \frac{1}{c}\bar{\mathbf{v}}\times\mathbf{H}\right) \qquad (13)$$

which is classical in terms of the electromagnetic field. This result has some importance as indicating that deflection methods of observing e/m will yield correct results when evaluated by the classical formulas.[6]

[1] Kennard, E. H., *Zeits. Physik*, **44**, 1927 (326–352).
[2] Darwin, C. G., *Proc. Roy. Soc.*, **117**, 1927 (258–297).
[3] Rabi, I. I., *Zeits. Phys.*, **49**, 1928 (507–511).
[4] Alexandrow, W., *Ibid.*, **56**, 1929 (818–837).
[5] Page, L., *Phys. Rev.*, **36**, 1930 (444–456).
[6] Cf. Eckart, C., *Ibid.*, **36**, 1930 (1514–1515); Kennard, E. H., *Ibid.*, **36**, 1930 (1667–1668).

REMARKS ON A PAPER: NOTE ON THE NATURE OF COSMIC RAYS, BY PAUL S. EPSTEIN

BY CARL STÖRMER

OSLO, NORWAY

Communicated December 5, 1930

In a recent paper, "Note on the Nature of Cosmic Rays," in the *Proc. Nat. Acad. Sci.* of the United States of America, October, 1930, Paul S. Epstein solved a problem concerning the motion of electrons in the field of an elementary magnet without being aware that this problem was solved as early as in the year 1904 in my first paper on the Polar Aurora.[1]

In fact, in the case of the Aurora, we meet exactly the same problem, viz., to find the regions of the earth which can be hit by electrons coming from very great distances. Thus Epstein's equation (11) corresponds to my equation (V, 1) and his equations (14) and (15) correspond to my equation (C, γ, k) on page 11 of my paper. A discussion of these equations was published in the above-mentioned paper and with further details in my Geneva paper[2] of 1907, which also contains the discussion made by Epstein, especially his inequality (18), where $1 + \sin^2\theta$ is erroneously printed instead of $1 + \sin^3\theta$.

A list of *approximate* maximum values of the angle $\theta_m^{(3)}$ will be found in my second Geneva paper[3] (1912), in the table at the beginning of §19. In that table I have found, e. g., the corresponding values of $\theta_m^{(3)}$ and $V = Ha$:

16°	291,000	and	30°	3,150,000
17°	368,000		31°	3,550,000
18°	460,000		32°	3,980,000

in agreement with Epstein's results.

A complete discussion of the regions of the earth hit by the cosmic rays, and also a study of the trajectories of these rays down to the earth can be made exactly on the same lines of research as those which I have developed in my mathematical theory of Aurora.[4,5,6,7]

In particular, the surface given in Epstein's paper

$$\rho_7 = \left(\frac{f}{A}\right)^{\frac{1}{3}} \frac{\sin^2\theta}{1 + \sqrt{1 + \sin^3\theta}}$$

is nothing else than the *toroidal surface* playing such an essential part in my explanation of the wireless echoes of long delay.[8,9]

[1] Carl Störmer, "Sur le mouvement d'un point matériel portant une charge d'électricité sous l'action d'un aimant élémentaire," *Videnskabsselskabets Skrifter*, 1904, Christiania.

[2] Carl Störmer, "Sur les trajectoires des corpuscules électrisés dans l'espace sous l'action du magnétisme terrestre avec application aux aurores boréales," *Archi Sci. Physiques et Naturelles*, Genève, 1907.

[3] Carl Störmer, "Sur les trajectoires des corpuscles electrisés dans l'espace sous l'action du magnétisme terrestre avec application aux aurores boréales, etc.," second memoire, Genève, 1911–1912.

For American readers I give the following references:

[4] Carl Störmer, "Corpuscular Theory of the Aurora Borealis," *Terrestrial Magnetism and Atmospheric Electricity*, March and September, 1917.

[5] Carl Störmer, "The Corpuscular Theory of Aurora Borealis," the book of the opening of the Rice Institute, **3**, Houston, Texas, 1917.

[6] Carl Störmer, "Twenty-five Years' Work on the Polar Aurora," *Terrestrial Magnetism and Atmospheric Electricity*, December, 1930.

Together with:

[7] Carl Störmer, "Periodische Elektronenbahnen im Felde eines Elementarmagneten und Ihre Anwendung auf Brüches Modellversuche und auf Eschenhagens Elementarwellen des Erdmagnetismus," *Zeits. Astrophysik*, November, 1930.

[8] Carl Störmer, "Kurzwellenechos die mehere Sekunden nach dem Hauptsignal eintreffen, und wie sie sich aus der Theorie des Polarlichtes erklären lassen," *Die Naturwissenschaften*, August 16, 1929, **17**, p. 643.

[9] Carl Störmer, "Do the Wireless Echoes of Long Delay Come from Space Outside the Moon's Orbit?" *Proc. Roy. Soc. Edinburgh*, **50**, Part II, No. 16, Edinburgh, 1930.

THE PRODUCTION OF HIGH SPEED CANAL RAYS WITHOUT THE USE OF HIGH VOLTAGES

By Ernest O. Lawrence and David H. Sloan

Department of Physics, University of California

Communicated December 15, 1930

Studies of collision processes involving electrons, ions and atoms have had an important part in the advances of our knowledge of the structure of matter and its interaction with radiation. Apart from studies with radioactive rays, such experimental investigations have been confined largely to impact processes involving kinetic energies of the order of magnitude of 100,000 volt-electrons and less. The investigation of the properties of high speed electrons and ions remains, therefore, a most fascinating field yet to be explored by the experimental physicist.

Though studies in the region of low energies have been so fruitful and have indicated the importance of knowledge of high energy collision processes, little progress in this direction has been made because of experimental difficulties. The chief difficulty concerns the production of the high speed corpuscles. The straightforward method of accelerating charged particles through requisite differences of potential presents great difficulties associated with the high electric fields necessarily involved. Apart from obvious troubles in obtaining high potentials with proper insulation, there is the problem of development of a vacuum tube suitable for the desired high voltages. In several laboratories[1,2,3,4] excellent progress has been made toward overcoming these difficulties, and indeed vacuum tubes have actually been developed which have been able to withstand potentials of the order of magnitude of one million volts. This technique, however, involves equipment so elaborate as practically to confine the work to a few laboratories in the world.

There is consequently a great need for the development of methods for the production of high speed particles not requiring high voltages and which therefore are free of the technical and practical difficulties that accompany high voltages. The present paper is an account of the first stage in such a development.

The fundamental principle of the methods we are developing was first demonstrated by Wideroe,[5] who obtained in a vacuum tube potassium ions having kinetic energies corresponding to twice the maximum voltage applied to the tube, which was about 25,000 volts. In the present work mercury ions having kinetic energies of over 200,000 volts have been obtained in a tube in which the maximum applied voltage was about 10,000 volts. As will become apparent below, much higher speed canal rays can be obtained using this technique.

A description of the experimental arrangement will make clear the important features of the method. A series of metal tubes arranged in line (labeled accelerator tubes in figure 1) are attached alternately to the inductance of a high frequency oscillatory circuit. A high frequency voltage applied in this manner produces at any instant electric fields between successive accelerator tubes of opposite direction and equal magnitude. If at one instant an ion finds itself between the first and second tubes it will be accelerated into the second tube, and if the time consumed in passing through this tube is equal to the half-period of the oscillator it will arrive between the second and third tubes with the field reversed in direction in such a manner that it will receive an additional acceleration on passing into the third tube. Each time the ion passes from the interior of one tube to the interior of the next it will gain energy proportional to the potential difference between the tubes. Hence if the tubes are made successively longer in the proper way (the square root of integers times the length of the first tube), for every frequency of the applied oscillations there will be a corresponding voltage such as will cause the ion to move up through the series of tubes in synchronism with the oscillating field, gaining between each pair of tubes an increase in kinetic energy corresponding to the applied potential difference.

This arrangement is not applicable to speeding up electrons to high kinetic energies because of the high speeds involved. Too long tubes would be required for obtainable frequencies of oscillation. Even for protons the requisite length of a series of tubes becomes hardly practicable, but in this instance we have overcome the difficulty by causing, with the aid of a magnetic field, the protons to traverse circular paths back and forth between semicircular hollow plates upon which a high frequency voltage is applied. This scheme is being developed at the present time in this laboratory and has been described already.[6] The method described in the present paper is quite ideal for the heavier ions. So far we have experimented with mercury ions only.

The first experimental tube contained eight accelerating tubes in line, thereby giving a theoretical amplification of eight-fold. It was eminently successful in demonstrating the passage of ions through the tubes, when accelerated by the proper voltage and frequency, and their emergence from the final tube with the theoretically expected velocity. Electrostatic deflection of the emergent beam showed that some of the ions had kinetic energies of as much as 90,000 volt-electrons when slightly in excess of 11,000 volts high frequency was applied to the accelerators. The ion currents in this instance were of the order of magnitude of 10^{-8} ampere.

The satisfactory performance of the first tube having been demonstrated, thirteen more accelerating tubes were next added to the original eight, making twenty-one accelerators in all. The constructional features of the

enlarged tube are shown in some detail in figure 1. The lower part contains a filament and a cylindrical anode with the end near the filament covered by a gauze. Mercury vapor at a pressure of slightly less than 10^{-3} mm. of mercury pressure was maintained in this region so that with the anode about 200 volts positive to the filament, a copious supply of mercury ions was formed in the interior of the anode. Next to the anode was placed a canal tube kept about at the filament potential. Positive ions were thus

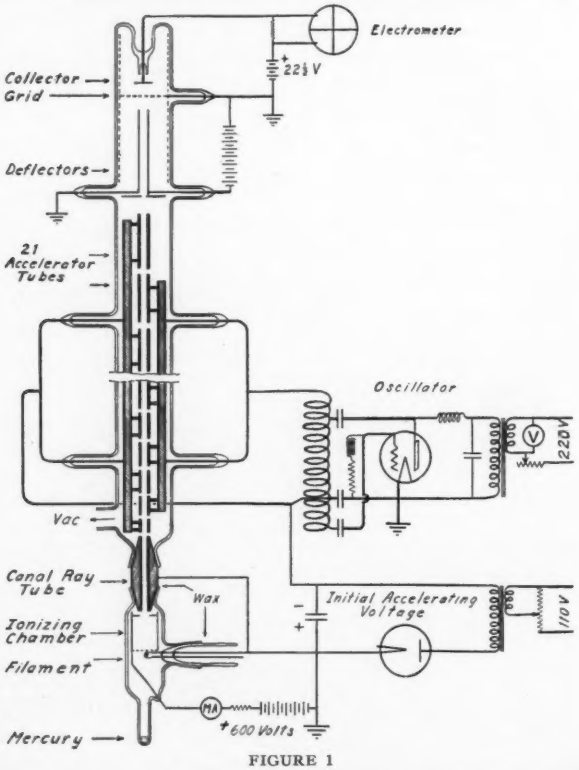


FIGURE 1

drawn from the interior of the anode up through the canal to the series of accelerating tubes. A steady d. c. potential of proper amount (2000 to 10,000 volts) accelerated the ions into the first of the accelerating tubes. The first accelerator had a length of 1.07 cm. Succeeding tubes were of increasing lengths, the last being 4.9 cm. long. All the tubes had the same diameters, 4 mm., and were made of nickel. The first eight were spaced from 1 to 2 mm. apart, while 4 mm. spacing was used on the last thirteen accelerators. Of course the lengths of the tubes were reduced below the

calculated values by the amounts of the spacing. Beyond the exit end of the tubes was placed a pair of deflecting plates 10 cm. long and 0.5 cm. apart, and further on was the collecting plate which was attached to an electrometer. The collector was screened by a surrounding gauze as was also the region of the deflecting plates.

The tube was first tried out using 118 meter oscillations on the accelerators. No accurate means was at hand for measuring the impressed high frequency voltages, but it was possible to make rough estimates from the voltage impressed on the oscillator tube. The oscillator tube was a 75-watt radiotron UX-852, the plate of which was supplied with a. c. excitation from a 110- to 10,000-volt transformer. It may not be out of place to remark that though the rated plate voltage of the UX-852 is 2000 volts, we found that the tube could withstand up to 15,000 volts. As the high frequency voltage impressed on the accelerators was raised, as indicated by the voltage impressed on the primary of the transformer, no current was picked up by the collector up to a certain point. Beyond this point the current rose rapidly with increasing voltage to a maximum, decreasing for higher voltages. This is shown by curve A of figure 2. Curve B shows a similar current variation with voltage for 85 meter oscillations. In both cases the ordinates record the currents to the collector corresponding to various high frequency voltages applied to the accelerators given by the abscissas. There are two maxima in evidence along curve B, one for a voltage near 3000 volts and another much higher maximum near 6000 volts. The higher maximum is interpreted as due to singly charged mercury ions, and the lower voltage maximum is attributed to doubly charged ions. A maximum produced by doubly charged ions when 118 meter oscillations were used was looked for later and found at the expected voltage.

Because no means was at hand to measure the high frequency voltages, it was assumed that such voltages were proportional to the plate voltages on the oscillator tube, which in turn were proportional to the voltages applied to the primary of the high voltage transformer. The voltage scale was calibrated on the assumption that the primary voltage giving the higher peak of curve B produced the calculated synchronizing voltage on the accelerators, i.e., 6200 volts. Such a procedure contains fairly large errors for the efficiency of the oscillator varies with wave-length, principally because of power loss to resonating circuits in the room, and the transformer voltage amplification ratio varies with the voltage on the primary. Since the power drawn from the transformer was less than 10 per cent of its capacity the latter source of error was not very large. Despite these sources of error this method of estimating the voltages is satisfactory to the extent that it will reveal the important features of the phenomena produced by the tube.

For example, doubly charged mercury ions require one half the voltage for acceleration in synchronism up through the tubes, which is seen to be the case in figure 2, curve B. The primary transformer voltages for the two peaks were 47 volts and 96 volts, respectively. The maximum ion current when 118 meters oscillations were used occurred for a primary voltage of 56 volts corresponding to a high frequency voltage on this scale of 3620 volts. The computed voltage for this wave-length is 3220 volts, which is represented on the diagram of figure 2 as a dotted line. The discrepancy is not large and undoubtedly is due to the different behavior of the oscillator at the two frequencies.

The wave-length of the oscillations was next reduced to 68 meters which required a high frequency voltage of 9750 volts on the accelerators to

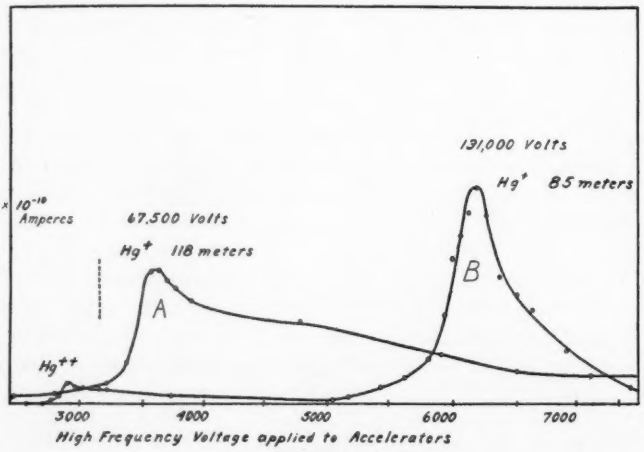


FIGURE 2

produce synchronism of the ions. As was expected, it was necessary to raise the primary voltage on the oscillator transformer to 120 volts in order to accelerate the ions through to the collector. In all the observations of ions arriving at the collector in these experiments, deflecting voltages were maintained on the deflecting plates thereby cutting out slow ions in the stream. A deflecting voltage of approximately 600 volts reduced the 118-meter ion current to a small value indicating that some of the ions had energies of about 60,000 or 70,000 volt-electrons. The whole available deflecting d. c. potential of 1200 volts was not able to reduce the 85 meter and 68 meter ion currents to any great extent indicating that in these cases the ions had energies in excess of 120,000 volts. Indeed, all observations of ion currents produced by 68 meter oscillations were made with the full 1200 volts on the deflecting plates. That 1200 volts should

fail to reduce the ion current in these two cases is to be expected, for the computed energies of the ions are 131,000 and 205,000 volt-electrons, respectively. There can be little doubt that such high speed ions were produced.

No particular efforts were made in these experiments to produce large ion currents. As is shown in figure 2, they were of the order of magnitude of 10^{-10} ampere. Unquestionably the magnitude of the currents can be greatly increased by better design of the canal ray source and enlarging the diameter of the accelerators. Accelerating and retarding potentials between the collector and surrounding grid showed that the measured currents consisted largely of secondary electrons ejected from the collector plate by the high speed ions. Rough observations, using the 205,000-volt ions, indicated that several secondary electrons were ejected by each impinging ion.

An attempt was made to apply about 15,000 volts to the accelerators at a correspondingly higher frequency, but it was found that the discharges occurred between the initial accelerating tubes which were close together. An effort was next made to increase the frequency and observe doubly charged ions coming through the tubes with 10,000 volts on the accelerators, thereby yielding ions with over 400,000 volt-electrons energy. But before this was accomplished the tube cracked and the effort was abandoned because it was evident that much more interesting results could be obtained by construction of another larger tube of better design. This is being done at the present time.

It is of interest to indicate to what equivalent voltages canal rays can be accelerated by this method. Twenty-five thousand volt high frequency oscillations at a wave-length of 43 meters applied to a tube containing 50 accelerators having an over-all length of 230 cm. is an entirely practicable laboratory arrangement which would yield 1,250,000-volt, singly charged mercury ions; 2,500,000-volt doubly charged ions would be obtainable from the same arrangement using 30 meters oscillations at the same voltage. Moreover, it appears quite feasible with the aid of water-cooled power oscillator tubes to impress 50,000 volts on the accelerators and thereby result in the production of 2,500,000-volt singly charged ions. Naturally because of the greater speeds involved, longer accelerator tubes would be required for lighter ions and hence greater difficulties would be encountered in accelerating them up to these extremely high energy values.

Thus, these initial experiments have demonstrated that high speed canal rays can be produced readily without the use of high potentials. Because the high speed ions are obtainable with quite modest laboratory equipment, the interesting and important studies of the properties of such high speed canal rays should receive wide attention.

[1] W. D. Coolidge, *J. Frank. Inst.*, **202**, 693 (1926).
[2] C. C. Lauritsen and R. D. Bennett, *Phys. Rev.*, **32**, 850 (1928).
[3] M. A. Tuve, L. R. Hafstad and O. Dahl, *Phys. Rev.*, **36**, No. 7, 1262 (1930).
[4] A. Brasch and F. Lange, *Naturwissenschaften*, **35**, 765 (Aug. 29, 1930).
[5] R. Wideroe, *Arch. Elektrot.*, **21**, 387 (1929).
[6] E. O. Lawrence and N. E. Edlefsen, *Science*, **72**, No. 1867, 376 (1930).

ON THE INTERPRETATION OF DIRAC'S α MATRICES

By G. Breit

Department of Physics, New York University

Communicated December 5, 1930

It is known that in a certain sense the matrices α_1, α_2, α_3 used by Dirac in his relativistic equation have the significance of velocities. It has been recently shown by Schroedinger[1] that in a certain sense the motion of a free electron can be decomposed into two parts. The part $\bar{x}_K$ represents the usual uniform motion of translation while ξ_K is a superposed motion of a vibratory character. Although Schroedinger's paper makes it clear just how the α_K may be thought of as instantaneous velocities it seems appropriate to point out some of the characteristic features of this interpretation. We wish to bring out that: (1) it is possible to have solutions of the general relativistic one electron problem which correspond to one of the α_K having definitely a given value say $+1$ or (-1) at one particular time and which imply that the velocity is $-c$ (or $+c$) in the direction x_K, (2) such solutions give a definite value of a α_K only at one time, (3) such solutions for a free electron involve in general a finite probability of states of negative energy, (4) in the special case of the momentum p_K becoming infinite the probability of states of negative energy may become zero.

In the notation of Dirac $\frac{dx_K}{dt} = -c\,\alpha_K$ and in order that $\alpha_1 = 1$ we must have $\psi_1 = \psi_4$, $\psi_2 = \psi_3$. The four Dirac equations are easily expressed in terms of

$$\chi_1 = \psi_1 - \psi_4,\ \chi_2 = \psi_2 - \psi_3,\ \chi_3 = \psi_1 + \psi_4,\ \chi_4 = \psi_2 + \psi_3 \quad (1)$$

and are

$$(p_0 - p_1)\chi_1 + (mc - ip_2)\chi_3 + p_3\chi_4 = 0 \quad (2.1)$$
$$(p_0 + p_1)\chi_3 + (mc + ip_2)\chi_1 - p_3\chi_2 = 0 \quad (2.2)$$
$$(p_0 + p_1)\chi_4 + (mc - ip_2)\chi_2 + p_3\chi_1 = 0 \quad (2.3)$$
$$(p_0 - p_1)\chi_2 + (mc + ip_2)\chi_4 - p_3\chi_3 = 0 \quad (2.4)$$

where $$p_0 = -\frac{h}{2\pi i}\frac{\partial}{c\partial t} + \frac{e}{c}A_0, \qquad p_K = \frac{h}{2\pi i}\frac{\partial}{\partial x_K} + \frac{e}{c}A_K \quad (2.5)$$

and the A_μ form the four potential. On account of (1)

$$\sum_1^4 \chi_\mu^* \chi_\mu = 2 \sum_1^4 \psi_\mu^* \psi_\mu. \tag{2.6}$$

For $\alpha_1 = 1$ we must have $\chi_1 = \chi_2 = 0$ for all (x_1, x_2, x_3) and given $t = t_0$. Then equations (2.2), (2.3) become

$$(p_0 + p_1)\chi_3 = (p_0 + p_1)\chi_4 = 0. \tag{3}$$

Using (2.5) we have on considering the complex conjugate of (3) and eliminating A_0, A_1

$$\left(\frac{\partial}{c\partial t} - \frac{\partial}{\partial x_1}\right)(\chi_3^* \chi_3) = \left(\frac{\partial}{c\partial t} - \frac{\partial}{\partial x_1}\right)(\chi_4^* \chi_4) = 0.$$

Since $\chi_1 = \chi_2 = 0$, we also have $\left(\frac{\partial}{c\partial t} - \frac{\partial}{\partial x_1}\right) \chi_1^* \chi_1 = \left(\frac{\partial}{c\partial t} - \frac{\partial}{\partial x_1}\right) \chi_2^* \chi_2 = 0,$

even though $\frac{\partial \chi_1}{\partial t} \neq 0$, hence using (2.6)

$$\left(\frac{\partial}{c\partial t} - \frac{\partial}{\partial x_1}\right) \sum_1^4 \psi_\mu^* \psi_\mu = 0. \tag{4}$$

Equation (4) is true at the particular $t = t_0$ and all (x_1, x_2, x_3). Combined with (2.6) it shows that for $t = t_0$ the charge density moves with the velocity c in the direction $-x_1$. The rotational invariance of Dirac's equation assures us that this result is independent of the choice of axes. We see, therefore, that it is possible to have solutions of Dirac's equation which correspond to the velocity being c at a given instant in a given direction (l_1, l_2, l_3). These solutions mean formally that $l_1\alpha_1 + l_2\alpha_2 + l_3\alpha_3$ is on a principal axis at that instant.

We have used so far equations (2.2) and (2.3). Equations (2.1) and (2.4) determine the rates of change of χ_1 and χ_2 at the time t_0. This rate of change may be shown to be distinct from zero so that it is impossible to have one of the α_K on its principal axes permanently. In fact, if $\frac{\partial \chi_1}{\partial t} = \frac{\partial \chi_2}{\partial t} = 0$, and $\chi_1 = \chi_2 = 0$, equations (2.1), (2.4) become on adding them to each other and also on subtracting:

$$\begin{aligned} mc\varphi_1 + (ip_2 + p_3)\varphi_2 &= 0 \\ mc\varphi_2 + (ip_2 - p_3)\varphi_1 &= 0 \end{aligned} \tag{5}$$

where $\varphi_2 = \chi_4 - \chi_3, \quad \varphi_1 = \chi_4 + \chi_3.$

The vector potentials may be eliminated from equations (5) and their complex conjugates. It then follows that

$$(2\pi mc/h)(\varphi_1^*\varphi_1 + \varphi_2^*\varphi_2) = \left(-\frac{\partial}{\partial x_2} + i\frac{\partial}{\partial x_3}\right)(\varphi_1^*\varphi_2) = \left(-\frac{\partial}{\partial x_2} - i\frac{\partial}{\partial x_3}\right)(\varphi_1\varphi_2^*). \quad (6)$$

Integrating both sides over the x_2, x_3 plane we obtain zero on the right, whenever φ_1, φ_2 vanish at infinity to a sufficiently high order. In order that the integral of the left side should have a finite contribution from the infinite region $\varphi_1^*\varphi_1 + \varphi_2^*\varphi_2$ must vanish at least as fast as $1/\rho^2$ where $\rho^2 = x_2^2 + x_3^2$. Under these conditions the integral of the right side vanishes and hence

$$\varphi_1 = \varphi_2 = 0 \quad (7)$$

everywhere. The above proof is not quite rigorous because a function may vanish as $1/\rho^2$ and faster and yet its derivative may not vanish so that the integrals of the right side of (6) may diverge. This has been pointed out to the writer by Prof. T. H. Gronwall who has been also kind enough to supply a mathematically rigorous proof applying with great generality to the case of vanishing vector potential. Since physically we cannot admit solutions implying divergence of the integrals of the right side of (6) we may say that *no admissible solutions of Dirac's equations exist which correspond to one α_k being permanently on one of its principal axes.*[2]

We thus conclude that from the purely formal point of view, i.e., taking Dirac's equation and using it according to the rules of quantum mechanics the matrices α_K have a general meaning of instantaneous velocities in the sense of equation (4), that it is possible to have solutions for which one α_K is instantaneously on a principal axis, but that it is impossible for the α_K to remain on a principal axis. The only definite value of a velocity component is therefore $\pm c$ but this definite value exists only instantaneously.

We now analyze the solutions $\chi_1 = \chi_2 = 0$ $(t = t_0)$ for a free electron. In order that such solutions should exist we must have a special type of Fourier analysis of the solutions. Every Fourier component has the form

$$\chi_\mu = a_\mu(\mathbf{p})\, exp\, [(2\pi\, i/h)(-Et + p_1x_1 + p_2x_2 + p_3x_3)]. \quad (8)$$

We substitute this into (2) and obtain

$$(E/c + p_1)a_3 = -(mc + ip_2)a_1 + p_3a_2 \quad (8')$$
$$(E/c + p_1)a_4 = -p_3a_1 - (mc - ip_2)a_2.$$

In order that at $t = 0$, $\chi_1 = \chi_2 = 0$ we must have

$$a_1(\mathbf{p}) + b_1(\mathbf{p}) = a_2(\mathbf{p}) + b_2(\mathbf{p}) = 0 \quad (9)$$

where b_μ refer to coefficients of exponentials with opposite signs of the energy E. The probability of an energy $|E|$ is proportional to

$$\sum_1^4 a_\mu^* a_\mu = (a_1^* a_1 + a_2^* a_2). \quad 2|E|/(|E| + cp_1) \tag{10}$$

as becomes clear on using (8'), and the probability of the energy $-|E|$ is similarly

$$\sum_1^4 b_\mu^* b_\mu = (b_1^* b_1 + b_2^* b_2). \quad 2|E|/(|E| - cp_1). \tag{11}$$

The probability $P(+)$ of the energy $+|E|$ is related to the probability $P(-)$ of the energy $-|E|$, by

$$P(+) = \frac{|E| - cp_1}{|E| + cp_1} \mathrm{P}(-). \tag{12}$$

The energy $|E|$ is connected with the momenta by

$$|E| = + \sqrt{m^2 c^2 + c^2 (p_1^2 + p_2^2 + p_3^2)}. \tag{13}$$

If p_1 is large and negative $P(+)/P(-)$ is small on account of (12). For moderate p_1 this ratio is of the order of 1. Thus if we wish to have $\alpha_1 = 1$ even instantaneously we must in general use states involving appreciable probabilities of negative energies. Only if the momentum is made infinite in the direction $-x_1$ can these states give a vanishingly small probability $P(-)$. In this case component $\tilde{x}_1$ of Schroedinger has a velocity $-c$.

We see therefore that even though formally the $-c\alpha_K$ are velocities, the absence of *negative energies* in the physical world and the apparently false prediction of these energies by Dirac's equation *necessitate considering only those* experimental *conditions* for the measurement of the α_K *which correspond* also *to a definitely infinite momentum* in the same direction as the velocity.

[1] E. Schroedinger, *Zitzungsberichte preuss. Akad.*, **24**, p. 418, 1930.

[2] Purely formal solutions becoming infinite exponentially at infinity may be constructed, however. Such solutions have no place in the general structure of quantum theory. If desired, one could interpret them as meaning that an electron at infinity (in the y direction) moves permanently with a velocity c in the perpendicular x direction.